

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

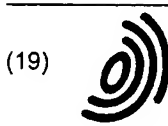
Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) **EP 0 682 710 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention  
of the grant of the patent:  
29.10.2003 Bulletin 2003/44

(51) Int Cl.7: **C12N 15/62**, C12N 1/19,  
C12N 1/20, C12N 11/16  
// (C12N15/62, 1:19,  
C12R1:465)

(21) Application number: **94908307.5**

(22) Date of filing: **10.02.1994**

(86) International application number:  
**PCT/EP94/00427**

(87) International publication number:  
**WO 94/018330 (18.08.1994 Gazette 1994/19)**

(54) **ISOLATION PROCESS USING IMMOBILIZED PROTEINS WITH SPECIFIC BINDING CAPACITIES**

ISOLIERUNGSVERFAHREN VERWENDEND IMMOBILISIERTE PROTEINE MIT SPEZIFISCHEN  
BINDUNGSKAPAZITÄTEN

PROCEDE D'ISOLATION UTILISANT DES PROTEINES IMMOBILISEES A CAPACITES DE  
FIXATION SPECIFIQUE

(84) Designated Contracting States:  
**AT BE CH DE DK ES FR GB GR IE IT LI NL PT SE**

(30) Priority: **10.02.1993 EP 93200350**

(43) Date of publication of application:  
**22.11.1995 Bulletin 1995/47**

(73) Proprietors:  
• **UNILEVER PLC**  
London EC4P 4BQ (GB)  
Designated Contracting States:  
**GB IE**  
• **UNILEVER N.V.**  
3013 AL Rotterdam (NL)  
Designated Contracting States:  
**AT BE CH DE DK ES FR GR IT LI NL PT SE**

(72) Inventors:  
• **FRENKEN, Leon Gerardus J., c/o Unilever  
Research**  
**NL-3133 AT Vlaardingen (NL)**

• **DE GEUS, Pieter**  
**NL-2991 KB Barendrecht (NL)**  
• **KLIS, Franciscus Maria**  
**NL-1025 KL Amsterdam (NL)**  
• **TOSCHKA, Holger York**  
**D-48734 Reken (DE)**  
• **VERRIPS, Cornelis Theodorus, c/o Unilever  
Research**  
**NL-3133 AT Vlaardingen (NL)**

(74) Representative: **Van Velzen, Maaïke Mathilde et al**  
**Unilever Patent Department**  
**Postbus 137**  
**3130 AC Vlaardingen (NL)**

(56) References cited:  
**EP-A- 0 244 221** **WO-A-92/04363**  
**WO-A-92/20805** **WO-A-94/01567**  
**US-A- 5 223 408**

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

**EP 0 682 710 B1**

## Description

## Background of the invention

5 [0001] The pharmaceutical, the fine chemicals and the food industry need a number of compounds that have to be isolated from complex mixtures such as extracts of animal or plant tissue, or fermentation broth. Often these isolation processes determine the price of the product.

[0002] Conventional isolation processes are not very specific and during the isolation processes the compound to be isolated is diluted considerably with the consequence that expensive steps for removing water or other solvents  
10 have to be applied.

[0003] For the isolation of some specific compounds affinity techniques are used. The advantage of these techniques is that the compounds bind very specifically to a certain ligand. However these ligands are quite often very expensive.

[0004] To avoid spillage of these expensive ligands they can be linked to an insoluble support. However, often linking the ligand is also expensive and, moreover, the functionality of the ligand is often affected negatively by such procedure.

15 [0005] So a need exists for developing cheap processes for preparing highly effective immobilized ligands.

[0006] The non-published co-pending international patent application WO 94/01567 (UNILEVER) discloses a method for immobilizing an enzyme or functional part thereof on the cell wall of a lower eukaryote, whereby the enzyme or functional part thereof is localized at the exterior of the cell wall.

[0007] WO 92/20805 discloses the production of heterologous proteins in Gram-positive bacteria wherein the heterologous proteins are bound to the surface of the bacteria by a cell wall spanning and membrane anchoring amino acid sequence.

[0008] J.E. Francisco et al, Proc. Natl. Acad. Sci. USA 89 (1992) 2713-2717, disclose a fusion protein obtained by the expression of a chimeric gene consisting of (i) the signal sequence and first nine N-terminal amino acids of the mature major *E. coli* lipoprotein, (ii) amino acids 46-159 of the outer membrane OmpA, and (iii) the complete mature  
25  $\beta$ -lactamase sequence.

[0009] The fusion protein had an enzymatically active  $\beta$ -lactamase and was found predominantly in the outer membrane. A substantial fraction (20-30%) of the  $\beta$ -lactamase domain of the protein was exposed on the external surface of *E. coli*.

[0010] In J. Bacteriol. 171 (1989) 4569-4576, the anchoring of  $\beta$ -galactosidase to the *E. coli* inner plasma membrane is disclosed. It is suggested that by combining a signal sequence at the N-terminus with the hydrophobic domain at the N-terminus, many proteins, although not  $\beta$ -galactosidase, could become associated with the outer aspect of the inner membrane or with the outer membrane. Such modified bacteria might have potential use as immobilized enzyme systems.

[0011] S.W. Hiebert et al, J. Cell Biology 107 (1988) 865-876, describe that the soluble cytoplasmic protein pyruvate kinase (PK) has been expressed at the cell surface of CV1 cells in a membrane-anchored form (APK). Truncated forms of the APK molecule, with up to 80% of the PK region of APK removed, can also be expressed at the cell surface.

[0012] In WO 89/07140 a tripartite DNA sequence is described that is so fused that the gene will upon expression in a eukaryotic cell (exemplified are COS cells) give rise to a correctly processed, appropriately folded, membrane bound version of the protein. Various anchor domains are mentioned.

40 [0013] C.C. Chen et al, J. Biol. Chem. 265 (1990) 3161-3167, describe that the predicted amino acid sequence reveals the presence of a proline-rich cell wall anchor region similar to that of other Gram-positive surface proteins.

[0014] Vijaya et al, Mol. Cell. Biol. 8 (1988) 1709-1714, describe the cell surface localization of hybrid proteins on the outer surface of CV-1 cells infected by recombinant viruses.

## 45 Summary of the invention

[0015] The invention is a process for carrying out an isolation process by using an immobilized binding protein or functional part thereof still capable of binding to a specific compound, wherein a medium containing said specific compound is contacted with a host cell according to the invention under conditions whereby a complex between said specific  
50 compound and said immobilized binding protein is formed, separating said complex from the medium originally containing said specific compound and, optionally, releasing said specific compound from said binding protein or functional part thereof.

[0016] Preferably, the host is selected from Gram-positive bacteria and fungi, which have a cell wall at the outside of the host cell, in contrast to Gram-negative bacteria and cells of higher eukaryotes such as animal cells and plant cells, which have a membrane at the outside of their cells. Suitable Gram-positive bacteria comprise lactic acid bacteria and bacteria belonging to the genera *Bacillus* and *Streptomyces*. Suitable fungi comprise yeasts belonging to the genera *Candida*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Pichia* and *Saccharomyces*, and moulds belonging to the genera *Aspergillus*, *Penicillium* and *Rhizopus*. In this specification the group of fungi comprises the group of yeasts  
55

and the group of moulds, which are also known as lower eukaryotes. In contrast to the cells in plants and animals, the group of bacteria and lower eukaryotes are also indicated in this specification as microorganisms.

[0017] Recombinant polynucleotides capable of being used in a method as described above, are polynucleotides comprising (i) a structural gene encoding a binding protein or a functional part thereof still having the specific binding capability, and (ii) at least part of a gene encoding an anchoring protein capable of anchoring in the cell wall of a Gram-positive bacterium or a fungus, said part of a gene encoding at least the anchoring part of said anchoring protein, which anchoring part is derivable from the C-terminal part of said anchoring protein.

[0018] The anchoring protein can be selected from  $\alpha$ -agglutinin,  $\alpha$ -agglutinin, FLO1, the Major Cell Wall Protein of a lower eukaryote, and proteinase of lactic acid bacteria. Preferably, such polynucleotide further comprises a nucleotide sequence encoding a signal peptide ensuring secretion of the expression product of the polynucleotide, which signal peptide can be derived from a protein selected from the  $\alpha$ -mating factor of yeast,  $\alpha$ -agglutinin of yeast, invertase of *Saccharomyces*, inulinase of *Kluyveromyces*,  $\alpha$ -amylase of *Bacillus*, and proteinase of lactic acid bacteria. The polynucleotide can be operably linked to a promoter, which is preferably an inducible promoter.

#### Brief description of the figures

##### [0019]

In Figure 1 the composition of pEMBL9-derived plasmid pUR4122 is indicated, the preparation of which is described in Example 1.

In Figure 2 the composition of plasmid pUR2741 is indicated, which is a derivative of published plasmid pUR2740, see Example 1.

In Figure 3 the composition of pEMBL9-derived plasmid pUR2968 is indicated. Its preparation is described in Example 1.

In Figure 4 the preparation of plasmid pUR4174 starting from plasmids pUR2741, pUR2968 and pUR4122 is indicated, as well as the preparation of plasmid pUR4175 starting from plasmids pSY16, pUR2968 and pUR4122. These preparations are described in Example 1.

In Figure 5 the composition of plasmid pUR2743.4 is indicated. Its preparation is described in Example 2. It contains the 714 bp *Pst*I-*Xho*I fragment given in SEQ ID NO: 12, which fragment encodes an scFv-TRAS fragment of anti-traseolide® antibody 02/01/01.

In Figure 6 the composition of plasmid pUR4178 is indicated. Its preparation is indicated in Example 2. It contains the above mentioned 714 bp *Pst*I-*Xho*I fragment given in SEQ ID NO: 12. This plasmid is suitable for the expression of a fusion protein between scFv-TRAS and  $\alpha$ AGG preceded by the invertase signal sequence (SUC2).

In Figure 7 the composition of plasmid pUR4179 is indicated. Its preparation is indicated in Example 2. It contains the above mentioned 714 bp *Pst*I-*Xho*I fragment given in SEQ ID NO: 12. This plasmid is suitable for the expression of a fusion protein between scFv-TRAS and  $\alpha$ AGG preceded by the prepro- $\alpha$ -mating factor signal sequence.

In Figure 8 a molecular design picture is given, showing the musk odour molecule traseolide® and a modified musk antigen, described in Example 3.

In Figure 9 the composition of plasmid pUR4177 is indicated. Its construction is described in Example 4. Plasmid pUR4177 contains the 734 bp *Eag*I-*Xho*I DNA fragment given in SEQ ID NO: 13 encoding the variable regions of the heavy and light chain fragments from the monoclonal antibody directed against the human chorionic gonadotropin (an scFv-HCG fragment) and is a 2  $\mu$ m-based vector suitable for production of the chimeric scFv HCG- $\alpha$ AGG fusion protein preceded by the invertase signal sequence and under the control of the GAL7 promoter.

In Figure 10 the composition of plasmid pUR4180 is indicated. Its preparation is indicated in Example 4. It contains the above mentioned 734 bp *Eag*I-*Xho*I DNA fragment given in SEQ ID NO: 13 and is a 2  $\mu$ m-based vector suitable for production of the chimeric scFv-HCG- $\alpha$ AGG fusion protein preceded by the prepro- $\alpha$ -mating factor signal sequence and under the control of the GAL7 promoter.

In Figure 11 the composition of plasmid pUR2990, a 2  $\mu$ m-based vector, is indicated, which is suggested in Example 5 as a starting vector for the preparation of plasmid pUR4196 (see Figure 12). Plasmid pUR2990 contains a DNA fragment encoding a chimeric lipase-FLO1 protein that will be anchored in the cell wall of a lower eukaryote and can catalyze lipid hydrolysis.

In Figure 12 the composition of plasmid pUR4196 is indicated. Its preparation is explained in Example 5. It contains a DNA fragment encoding a chimeric protein comprising the scFv-HCG followed by the C-terminal part of the FLO1-protein, and is a vector suitable for the production of a chimeric protein anchored in the cell wall of the host organism and can bind HCG.

In Figure 13 the composition of plasmid pUR2985 is indicated. Its preparation is described in Example 6. It contains a *choB* gene coding for the mature part of the cholesterol oxidase (EC 1.1.3.6) obtained via PCR techniques from the chromosome of *Brevibacterium sterolicum*.

In **Figure 14** the composition of plasmid pUR2987 is indicated. Its preparation from plasmid pUR2985 is described in Example 6. It contains a DNA sequence comprising the *choB* gene coding for the mature part of the cholesterol oxidase preceded by DNA encoding the prepro- $\alpha$ -mating factor signal sequence and followed by DNA encoding the C-terminal part of  $\alpha$ -agglutinin.

In **Figure 15** the composition of the published plasmid pGKV550 is indicated. It is described in Example 7 and contains the complete cell wall proteinase operon of *Lactococcus lactis* subsp. *cremoris* Wg2, including the promoter, the ribosome binding site and the prtP gene.

In **Figure 16** the composition of plasmid pUR2988 is indicated. Its preparation is described in Example 7. It is anticipated that this plasmid can be used for preparing a further plasmid pUR2989, which after introduction in a lactic acid bacterium will be responsible for producing a chimeric protein that will be anchored at the outer surface of the lactic acid bacterium and is capable of binding cholesterol.

In **Figure 17** the composition of plasmid pUR2993 is indicated. Its preparation is described in Example 8. It is anticipated that this plasmid can be used for transforming yeast cells that can bind a human epidermal growth factor (EGF) through an anchored chimeric protein containing an EGF receptor.

In **Figure 18** the composition of plasmids pUR4482 and 4483 is indicated. Their preparation is described in Example 9. Plasmid pUR4482 is a yeast episomal expression plasmid for expression of a fusion protein with the invertase signal sequence, the CH<sub>v</sub>09 variable region, the Myc-tail, and the "X-P-X-P" Hinge region of a camel antibody, and the  $\alpha$ -agglutinin cell wall anchor region. Plasmid pUR4483 differs from pUR4482 in that it does not contain the "X-P-X-P" Hinge region.

In **Figure 19** immunofluorescent labelling (anti-Myc antibody) of SU10 cells in the exponential phase ( $OD_{530} = 0.5$ ) expressing the genes of camel antibodies present on plasmids pUR4424, pUR4482 and pUR4483 is shown. Ph = phase contrast, Fl = fluorescence.

In **Figure 20** immunofluorescent labelling (anti-human IgG antibody) of SU10 cells in the exponential phase ( $OD_{530}=0.5$ ) expressing the genes of camel antibodies present on plasmids pUR4424, pUR4482 and pUR4483 is shown.

Ph = phase contrast, Fl = fluorescence.

#### Abbreviations used in the Figures:

#### [0020]

$\alpha$ -gal:	gene encoding guar $\alpha$ -galactosidase
AG- $\alpha$ 1/AG $\alpha$ 1:	gene expressing $\alpha$ -agglutinin from <i>S. cerevisiae</i>
AG $\alpha$ 1 cds/ $\alpha$ -AGG:	coding sequence of $\alpha$ -agglutinin
Amp/amp <sup>r</sup> :	$\beta$ -lactamase resistance gene
CH <sub>v</sub> 09:	camel heavy chain variable 09 fragment
EmR:	erythromycin resistance gene
f1:	phage f1 replication sequence
FLO1/FLO (C-part):	C-terminal part of FLO1 coding sequence of flocculation protein
Hinge:	Camel "X-P-X-P" Hinge region, see Example 9
LEU2:	LEU2 gene
LEU2d/Leu2d:	truncated LEU2 gene
Leu 2d cs:	coding sequence LEU2d gene
MycT:	camel Myc-tail
Ori MB1:	origin of replication MB1 derived from <i>E. coli</i> plasmid
Pga17/pGAL7:	GAL7 promoter
Tpgk:	terminator of the phosphoglyceratekinase gene
pp $\alpha$ -MF/MF $\alpha$ 1ss:	prepro-part of $\alpha$ -mating factor (= signal sequence)
repA:	gene encoding the repA protein required for replication (Fig. 15/16).
ScFv (V <sub>H</sub> -V <sub>L</sub> ):	single chain antibody fragment containing V <sub>H</sub> and V <sub>L</sub> chains
ss:	signal sequence
SUC2:	invertase signal sequence
2u/2 micron:	2 $\mu$ m sequence

#### Detailed description of the invention

[0021] The present invention relates to the isolation of valuable compounds from complex mixtures by making use of immobilized ligands. The immobilized ligands can be proteins obtainable via genetic engineering and can consist

of two parts, namely both an anchoring protein or functional part thereof and a binding protein or functional part thereof.

[0022] The anchoring protein sticks into cell walls of microorganisms, preferably lower eukaryotes, e.g. yeasts and moulds. Often this type of proteins has a long C-terminal part that anchors it in the cell wall. These C-terminal parts have very special amino acid sequences. A typical example is anchoring via C-terminal sequences of proteins enriched in proline, see Kok (1990).

[0023] The C-terminal part of these anchoring proteins can contain a substantial number of potential serine and threonine glycosylation sites. O-glycosylation of these sites gives a rod-like conformation to the C-terminal part of these proteins.

In the case of anchored manno-proteins they seem to be linked to the glucan in the cell wall of lower eukaryotes, as they cannot be extracted from the cell wall with sodium dodecyl sulphate (SDS), but can be liberated by glucanase treatment, see our co-pending patent application WO-94/01567 (UNILEVER) published 20 January 1994 and Schreuder c.s. (1993), both being published after the claimed priority date. WO-A-94/01567 discloses the use of anchored enzymes in an enzymatic process. Another mechanism to anchor proteins at the outer side of a cell is to make use of the property that a protein containing a glycosyl-phosphatidyl-inositol (GPI) group anchors via this GPI group to the cell surface, see Conzelmann c.s. (1990).

[0024] The binding protein is so called, because it ligates or binds to the specific compound to be isolated. If the N-terminal part of the anchoring protein is sufficiently capable of binding to a specific compound, the anchoring protein itself can be used in a process for isolating that specific compound. Suitable examples of a binding protein comprise an antibody, an antibody fragment, a combination of antibody fragments, a receptor protein, an inactivated enzyme still capable of binding the corresponding substrate, and a peptide obtained via Applied Molecular Evolution, see Lewin (1990), as well as a part of any of these proteinaceous substances still capable of binding to the specific compound to be isolated. All these binding proteins are characterized by specific recognition of the compounds or group of related compounds to be isolated. The binding rate and release rate, and therefore the binding constant between the specific compound to be isolated and the binding protein, can be regulated either by changing the composition of the liquid extract in which the compound is present or, preferably, by changing the binding protein by protein engineering.

[0025] The gene coding for the chimeric protein comprising both the binding protein and the anchoring protein (or functional parts thereof) can be placed under control of a constitutive, inducible or derepressible promoter and will generally be preceded by a DNA fragment encoding a signal sequence ensuring efficient secretion of the chimeric protein. Upon secretion the chimeric protein will be anchored in the cell wall of the microorganisms, thereby covering the surface of the microorganisms with the chimeric protein. These microorganisms can be obtained in normal fermentation processes and their isolation is a cheap process, when physical separation processes are used, e.g. centrifugation or membrane filtration.

[0026] After washing, the isolated microorganisms can be added to liquid extracts containing the valuable specific compound or compounds. After some time the equilibrium between the bound and free specific compound(s) will be reached and the microorganisms to which the specific compound or group of related compounds is bound can be separated from the extract by simple physical techniques. Alternatively, the microorganisms covered with ligands can be brought on a support material and subsequently this coated support material can be used in a column.

[0027] The liquid extract containing the specific compound or compounds of interest can be added to the column and afterwards the compound(s) can be released from the ligand by changing the composition of the eluting liquid or the temperature or both. A skilled person will recognize that in addition to these two possibilities other modifications can be used for effecting the binding of the specific compound and the ligand, their subsequent isolation and/or the release of the specific compound(s).

[0028] In particular the invention relates to chimeric proteins that are bound to the cell wall of lower eukaryotes. Suitable lower eukaryotes comprise yeasts, e.g. *Candida*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Pichia* and *Saccharomyces*, and moulds e.g. *Aspergillus*, *Penicillium* and *Rhizopus*. For some applications prokaryotes are also applicable, especially Gram-positive bacteria, examples of which include lactic acid bacteria, and bacteria belonging to the genera *Bacillus* and *Streptomyces*.

[0029] For lower eukaryotes the present invention provides genes encoding chimeric proteins consisting of:

- a. a DNA sequence encoding a signal sequence functional in a lower eukaryotic host, e.g. derived from a yeast protein including the  $\alpha$ -mating factor, invertase,  $\alpha$ -agglutinin, inulinase or derived from a mould protein e.g. xylanase;
- b. a structural gene encoding a C-terminal part of a cell wall protein preceded by a structural gene encoding a protein, that is capable of binding to the specific compound or group of compounds of interest, examples of which include

- an antibody,
- a single chain antibody fragment (scFv; see Bird and Webb Walker (1991),

- a variable region of the heavy chain ( $V_H$ ) or a variable region of the light chain ( $V_L$ ) of an antibody or that part of such variable region still containing one to three of the complementarity determining regions (CDRs),
- an agonist-recognizing part of a receptor protein or a part thereof still capable of binding the agonist,
- a catalytically inactivated enzyme, or a fragment of such enzyme still containing a substrate binding site of the enzyme,
- specific lipid binding proteins or parts of these proteins still containing the lipid binding site(s), see Ossendorp (1992), and
- a peptide that has been obtained via Applied Molecular Evolution, see Lewin (1990).

[0030] All expression products of these genes are characterized in that they consists of a signal sequence and both a protein part, that is capable of binding to the compound(s) to be isolated, and a C-terminus of a typically cell wall bound protein, examples of the latter including  $\alpha$ -agglutinin, see Lipke c.s. (1989), a-agglutinin, see Roy c.s. (1991), FLO1 (see Example 5 and SEQ ID NO: 14) and the Major Cell Wall Protein of lower eukaryotes, which C-terminus is capable of anchoring the expression product in the cell wall of the lower eukaryote host organism.

[0031] The expression of these genes encoding chimeric proteins can be under control of a constitutive promoter, but an inducible promoter is preferred, suitable examples of which include the GAL7 promoter from *Saccharomyces*, the inulinase promoter from *Kluyveromyces*, the methanol-oxidase promoter from *Hansenula*, and the xylanase promoter of *Aspergillus*. Preferably the constructs are made in such a way that the new genetic information is integrated in a stable way in the chromosome of the host cell, see e.g. WO-91/00920 (UNILEVER).

[0032] The lower eukaryotes transformed with the above mentioned genes can be grown in normal fermentation, continuous fermentation, or fed batch fermentation processes.

[0033] The selection of a suitable process for growing the microorganism will depend on the construction of the gene and the promoter used, and on the desired purity of the cells after the physical separation procedure(s).

[0034] For bacteria the present invention deals with genes encoding chimeric proteins consisting of:

- a DNA sequence encoding a signal sequence functional in the specific bacterium, e.g. derived from a *Bacillus*  $\alpha$ -amylase, a *Bacillus subtilis* subtilisin, or a *Lactococcus lactis* subsp. *cremoris* proteinase;
- a structural gene encoding a C-terminal part of a cell wall protein preceded by a structural gene encoding a protein capable of binding to the specific compound or group of compounds of interest, examples of which are given above for a lower eukaryote.

[0035] All expression products of these genes are characterized in that they consist of a signal sequence and both a protein part, that is capable of binding to the specific compound or specific group of compounds to be isolated, and a C-terminus of a typically cell wall-bound protein such as the proteinase of *Lactococcus lactis* subsp. *cremoris* strain Wg2, see Kok c.s. (1988) and Kok (1990), the C-terminus of which is capable of anchoring the expression product in the cell wall of the host bacterium.

[0036] The invention is illustrated with the following Examples without being limited thereto. First the endonuclease restriction sites mentioned in the Examples are given.

<b>BstEII</b>	<b>G GTNACC</b> <b>CCANTG G</b>	<b>ClaI</b>	<b>AT CGAT</b> <b>TAGC TA</b>	<b>EagI</b>	<b>C GGCCG</b> <b>GCCGG C</b>
<b>EcoRI</b>	<b>G AATTC</b> <b>CTTAA G</b>	<b>HindIII</b>	<b>A AGCTT</b> <b>TTCGA A</b>	<b>NheI</b>	<b>G CTAGC</b> <b>CGATC G</b>
<b>NotI</b>	<b>GC GGCCGC</b> <b>CGCCGG CG</b>	<b>NruI</b>	<b>TCG CGA</b> <b>AGC GCT</b>	<b>PstI</b>	<b>CTGCA G</b> <b>G ACGTC</b>
<b>SacI</b>	<b>GAGCT C</b> <b>C TCGAG</b>	<b>SalI</b>	<b>G TCGAC</b> <b>CAGCT G</b>	<b>XhoI</b>	<b>C TCGAG</b> <b>GAGCT C</b>

**Example 1. Construction of a gene encoding a chimeric protein that will be anchored in the cell wall of a lower eukaryote and is able to bind with high specificity lysozyme from a complex mixture.**

[0037] Lysozyme is an anti-microbial enzyme with a number of applications in the pharmaceutical and food industries. Several sources of lysozyme are known, e.g. egg yolk or a fermentation broth containing a microorganism producing lysozyme. Monoclonal antibodies have been raised against lysozyme, see Ward c.s. (1989), and the mRNA's encoding the light and heavy chains of such antibodies have been isolated from the hybridoma cells and used as template for the synthesis of cDNA using reverse transcriptase. Starting from the plasmids as described by Ward c.s. (1989), we constructed a pEMBL-derived plasmid, designated pUR4122, in which the multiple cloning site of the pEMBL-vector, ranging from the *EcoRI* to the *HindIII* site, was replaced by a 231 bp DNA fragment, whose nucleotide sequence is given in SEQ ID NO: 1 and has an *EcoRI* site (GAATTC) at nucleotides 1-6, a *PstI* site (CTGCAG) at nucleotides 105-110, a *BstEII* site (GGTCACC) at nucleotides 122-128, a *XhoI* site (CTCGAG) at nucleotides 207-212, and a *HindIII* site (AAGCTT) at nucleotides 226-231.

#### 15 Construction of pUR4122

[0038] Plasmid pEMBL9, see Dente c.s. (1983), was digested with *EcoRI* and *HindIII* and the resulting large fragment was ligated with the double stranded synthetic DNA fragment given in SEQ ID NO: 1. For the successive ligation of DNA fragments, which finally form the coding sequence of a single chain antibody fragment for lysozyme, the following elements were combined in the 231 bp DNA fragment (SEQ ID NO: 1) inserted into the pEMBL-9 vector: the 3' part of the GAL7 promoter, the invertase signal sequence (SUC2), a *PstI* restriction site, a *BstEII* restriction site, a sequence encoding the (GGGGS)<sub>3</sub> peptide linker connecting the V<sub>H</sub> and V<sub>L</sub> fragments, a *SacI* restriction site, a *XhoI* restriction site and a *HindIII* restriction site, resulting in plasmid pUR4119. To obtain the in frame fusion between V<sub>H</sub> and the GGGGS-linker plasmid pSW1-VHD1.3-VKD1.3-TAG1, see Ward c.s. (1989), was digested with *PstI* and *BstEII* and a DNA fragment of 0.35 kbp was ligated in the correspondingly digested pUR4119 resulting in plasmid pUR4119A. Subsequently the plasmid pSW1-VHD1.3-VKD1.3-TAG1 was digested with *SacI* and *XhoI* and this fragment containing the coding part of V<sub>L</sub> was finally ligated into the *SacI/XhoI* sites of pUR4119A, resulting in plasmid pUR4122 (see Figure 1).

#### 30 Construction of pUR4174, see Figure 4

[0039] To obtain *S. cerevisiae* episomal expression plasmids containing DNA encoding a cell wall anchor derived from the C-terminal part of  $\alpha$ -agglutinin, plasmid pUR2741 (see Figure 2) was selected as starting vector. Basically, this plasmid is a derivative of pUR2740, which is a derivative of plasmid pUR2730 as described in WO-91/19782 (UNILEVER) and by Verbakel (1991). The preparation of pUR2730 is clearly described in Example 9 of EP-A1-0255153 (UNILEVER). Plasmid pUR2741 differs from plasmid pUR2740 in that the *EagI* restriction site within the remaining part of the already inactive *tet* resistance gene was deleted through *NruI/SalI* digestion. The *SalI* site was filled in prior to religation.

[0040] After digesting pUR4122 with *SacI* (partially) and *HindIII*, the approximately 800 bp fragment was isolated and cloned into the pUR2741 vector fragment, which was obtained after digestion of pUR2741 with the same enzymes. The resulting plasmid was named pUR4125.

[0041] A plasmid named pUR2968 (see Figure 3) was made by (1) digesting with *HindIII* the *Ag $\alpha$ 1*-containing plasmid pL $\alpha$ 21 published by Lipke c.s. (1989), (2) isolating an about 6.1 kbp fragment and (3) ligating that fragment with *HindIII*-treated pEMBL9, so that the 6.1 kbp fragment was introduced into the *HindIII* site present in the multiple cloning site of the pEMBL9 vector.

[0042] Plasmid pUR4125 was digested with *XhoI* and *HindIII* and the about 8 kbp fragment was ligated with the approximately 1.4 kbp *NheI-HindIII* fragment of pUR2968, using *XhoI/NheI* adapters having the following sequence:

<i>XhoI</i>	<i>NheI</i>	
5' - <u>TC GAG</u> ATC AAA GGC GGA TCT <u>G</u> -3'		= SEQ ID NO: 2
3' - <u>C</u> TAG TTT CCG CCT AGA <u>CGATC</u> -5'		= SEQ ID NO: 3.

[0043] The plasmid resulting from the ligation of the appropriate parts of plasmids pUR2968, pUR4125 and *XhoI/NheI* adapters, was designated pUR4174 and encodes a chimeric fusion protein at the amino terminus consisting of the invertase signal (pre) peptide, followed by the scFv-LYS polypeptide and, finally, the C-terminal part of  $\alpha$ -agglutinin (see Figure 4).



Construction of pUR4175, see Figure 4

**[0044]** Upon digesting pUR4122 (see above) with *Pst*I and *Hind*III, the approximately 700 bp fragment was isolated and ligated into a vector fragment of plasmid pSY16, see Harmsen c.s. (1993), which was digested with *Eag*I and *Hind*III and using *Eag*I-*Pst*I adapters, having the following sequence:

EagI PstI  
5'-G GCC GCC CAG GTG CAG CTG CA-3' = SEQ ID NO: 4  
3'-CGG GTC CAC GTC G -5' = SEQ ID NO: 5

**[0045]** The resulting plasmid, named pUR4132, was digested with *Xho*I and *Hind*III and ligated with the approximately 1.4 kbp *Nhe*I-*Hind*III fragment of pUR2968 (see above), using *Xho*I/*Nhe*I adapters as described above, resulting in pUR4175 (see Figure 4). This plasmid contains a gene encoding a chimeric protein consisting of the  $\alpha$ -mating factor prepro-peptide, followed by the scFv-LYS polypeptide and, finally, the C-terminal part of  $\alpha$ -agglutinin.

**Example 2. Construction of genes encoding a series of homologous chimeric proteins that will be anchored in the cell wall of a lower eukaryote and are able to bind with high specificities the musk fragrance trascolide® from a complex mixture.**

**[0046]** The isolation of RNA from the hybridoma cell lines, the preparation of cDNA and amplification of gene fragments encoding the variable regions of antibodies by PCR was performed according to standard procedures known from the literature, see e.g. **Orlandi c.s. (1989)**. For the PCR amplification different oligonucleotide primers have been used.

[0047] For the heavy chain fragment:

A: AGG TSM ARC TGC AGS AGT CWG G = SEQ ID NO: 6  
PstI.

in which S is C or G, M is A or C, R is A or G, and W is A or T,  
and

B: TGA GGA GAC GGT GAC CGT GGT CCC TTG GCC CC  
BstEII = SEQ ID NO: 7.

For the light chain fragment (Kappa):

C: GAC ATT GAG CTC ACC CAG TCT CCA = SEQ ID NO: 8,  
SacI

and

D: GTT TGA TCT CGA GCT TGG TCC C = SEQ ID NO: 9.  
XhoI

### Construction of pUR4143

**[0048]** To simplify future construction work an *EagI* restriction site was introduced in pUR4122 (see above), at the junction between the invertase signal sequence and the scFv-LYS. This was achieved by replacing the about 110 bp *EcoRI-PstI* fragment within the synthetic fragment given in SEQ ID NO: 1 by synthetic adapters with the following sequence:

*EcoRI* *PstI*  
AATTCGGCCGTTTCAGGTGCAGCTGCA  
GCCGGCAAGTCCACGTCG

= SEQ ID NO: 10  
 = SEQ ID NO: 11.

5

[0049] The resulting plasmid was designated pUR4122.1: a construction vector for single chain Fv assembly in frame behind an *EagI* site for expression behind either the prepro- $\alpha$ -mating factor sequence or the SUC2 invertase signal sequence.

10

[0050] After digesting the heavy chain PCR fragment with *PstI* and *BstEII*, two fragments were obtained: a *PstI* fragment of about 230 bp and a *PstI/BstEII* fragment of about 110 bp. The latter fragment was cloned into vector pUR4122.1, which was digested with *PstI* and *BstEII*. The newly obtained plasmid (pUR4122.2) was digested with *SacI* and *XhoI*, after which the light chain PCR fragment (digested with the same restriction enzymes) was cloned into the vector, resulting in pUR4122.3. This plasmid was digested with *PstI*, after which the above described about 230 bp *PstI* fragment was cloned into the plasmid vector, resulting in a plasmid called pUR4143. Two orientations are possible, but selection can be made by restriction analysis, as usual. Instead of the scFv-LYS gene originally present in pUR4122, this new plasmid pUR4143 contains a gene encoding an scFv-TRAS fragment of anti-traseolide antibody 02/01/01 (for the nucleotide sequence of the 714 bp *PstI-XhoI* fragment see SEQ ID NO: 12).

15

#### Construction of pUR4178 and pUR4179.

[0051] After digesting pUR4143 with *EagI* and with *HindIII*, an about 715 bp fragment can be isolated. Subsequently, this fragment can be cloned into the vector backbone fragments of pUR2741 and pUR4175, that were digested with the same restriction enzymes. In the case of pUR2741, this resulted in plasmid pUR2743.4 (see Figure 5). This plasmid can subsequently be cleaved with *XhoI* and *HindIII* and ligated with the about 8 kbp *XhoI-HindIII* fragment of pUR4174, resulting in pUR4178 (see Figure 6).

25

[0052] In the situation where pUR4175 was used as a starting vector, the resulting plasmid was designated pUR4179 (see Figure 7).

30

Both plasmids, pUR4178 and pUR4179 were introduced into *S. cerevisiae*.

**Example 3. The modification of the binding parts of the chimeric protein that can bind traseolide® in order to improve the binding or release of traseolide® under certain conditions.**

[0053] Modification of binding properties of antibodies during the immune response is a well known immunological phenomenon originating from the fine tuning of complementarity determining sequences in the antibody's binding region to the antigen's molecular properties. This phenomenon can be mimicked *in vitro* by adjusting the antigen binding regions of antibody fragments based on molecular models of these regions in contact with the antigen.

35

[0054] One such example consists of protein engineering the antimusk antibody M02/01/01 to a stronger binding variant M020501i.

40

[0055] First, a molecular model of M02/01/01 variable fragment (Fv) was constructed by homology modelling, using the coordinates of the anti-lysozyme antibody HYHEL-10 as a template (Brookhaven Protein Data Bank entry: 3HFM). This model was refined using Molecular Mechanics and Molecular Dynamics methods from within the Biosym program DISCOVER, on a Silicon Graphics 4D240 workstation. Secondly, the binding site of the resulting Fv was mapped by visually docking the musk antigen into the CDR region, followed by a refinement using molecular dynamics again. Upon inspection of the resulting model for packing efficiency (van der Waals contact areas), it was concluded that substitution of ALA H96 by VAL would increase the (hydrophobic) contact area between the ligand and Fv, and consequently lead to a stronger interaction (see Figure 8).

45

[0056] When this mutation is introduced into M02/01/01, the cDNA-derived scFv from Example 2, the result will be Fv M020501i; a variant with an increased affinity of at least a factor of 5 can be expected, and the increased affinity could be measured using fluorescence titration of the Fv with the musk odour molecule.

50

**Example 4. Construction of a gene encoding a chimeric protein that will be anchored in the cell wall of lower eukaryote and is able to bind hormones such as HCG.**

55

[0057] Gene fragments, encoding the variable regions of the heavy and light chain fragments from the monoclonal antibody directed against the human chorionic gonadotropin were obtained from a hybridoma cell line in a similar way as described in Example 2.

[0058] Subsequently, these HCG V<sub>11</sub> and V<sub>1</sub> gene fragments were cloned into plasmid pUR4143 by replacing the corresponding *Pst*I-*Bst*EII and *Sac*I-*Xho*I gene fragments, resulting in plasmid pUR4146.

[0059] Similar to the method described in Example 2, the 734 bp *Eag*I-*Xho*I fragment (nucleotide sequence given in SEQ ID NO: 13) encoding the variable regions of the heavy and light chain fragments from the monoclonal antibody directed against the human chorionic gonadotropin (an scFv-HCG fragment) was isolated from pUR4146 and was introduced into the vector backbone fragment of pUR4178 (see Example 2) and will be introduced into the vector backbone fragment of pUR4175 (see Example 1), both digested with the same restriction enzymes. The resulting plasmids pUR4177 (see Figure 9) was, and pUR4180 (see Figure 10) will be, introduced into *S. cerevisiae* strain SU10.

**Example 5. Construction of a gene encoding a chimeric scFv-FLO1 protein that will be anchored in the cell wall of lower eukaryote and is able to bind hormones such as HCG.**

[0060] One of the genes associated with the flocculation phenotype in *S. cerevisiae* is the FLO1 gene. The DNA sequence of a clone containing major parts of the FLO1 gene has been determined, see SEQ ID NO: 14 giving 2685 bp of the FLO1 gene. The cloned fragment appeared to be approximately 2 kb shorter than the genomic copy as judged from Southern and Northern hybridizations, but encloses both ends of the FLO1 gene. Analysis of the DNA sequence data indicates that the putative protein contains at the N-terminus a hydrophobic region which confirms a signal sequence for secretion, a hydrophobic C-terminus that might function as a signal for the attachment of a GPI-anchor and many glycosylation sites, especially in the C-terminus, with 46.6% serine and threonine in the arbitrarily defined C-terminus (aa 271-894). Hence, it is likely that the FLO1 gene product is located in an orientated fashion in the yeast cell wall and may be directly involved in the process of interaction with neighbouring cells.

[0061] The cloned FLO1 sequence might therefore be suitable for the immobilization of proteins or peptides on the cell surface by a different type of cell wall anchor.

[0062] For the production of a chimeric protein comprising the scFv-HCG followed by the C-terminal part of the FLO1-protein, plasmid pUR2990 (see Figure 11) can be used as a starting vector. The preparation of episomal plasmid pUR2990 was described in our co-pending patent application WO-94/01567 (UNILEVER) published on 20 January 1994, i.e. during the priority year. Plasmid pUR2990 comprises the chimeric gene consisting of the gene encoding the *Humicola* lipase and a gene encoding the putative C-terminal cell wall anchor domain of the FLO1 gene product, the chimeric gene being preceded by the invertase signal sequence (SUC2) and the GAL7 promoter; further the plasmid comprises the yeast 2  $\mu$ m sequence, the defective Leu2 promoter described by Eckard and Hollenberg (1983), and the Leu2 gene, see Roy c.s. (1991). Plasmid pUR4146, described in Example 4, can be digested with *Pst*I and *Xho*I, and the about 0.7 kbp *Pst*I-*Xho*I fragment containing the scFv-HCG coding sequence can be isolated. For the in frame fusion of this DNA sequence between the C-terminal FLO1 part and the SUC2 signal sequence, the fragment can be directly ligated with the 9,3 kbp *Eag*I/*Nhe*I (partial) backbone of plasmid pUR2990, resulting in plasmid pUR4196 (see Figure 12). This plasmid will comprise an additional triplet encoding Ala at the transition between the SUC2 signal sequence and the start of the scFv-HCG, and a E-I-K-G-G amino acid sequence in front of the first amino acid (Ser) of the C part of FLO1 protein.

[0063] If in the previous Examples 1-5 the level of exposed antibody fragments is too low, the production level can be increased by mutagenesis of the frame work regions of the antibody fragment. This can be done in a site directed way or by (targeted) random mutagenesis, using techniques described in the literature.

**Example 6. Construction of a gene encoding a chimeric protein that will be anchored in the cell wall of a lower eukaryote and is able to bind cholesterol.**

[0064] In the literature two DNA sequences for cholesterol oxidase are described, the *choB* gene from *Brevibacterium sterolicum*, see Ohta c.s. (1991) and the *choA* gene from *Streptomyces* sp. SA-COO, see Ishizaka c.s. (1989). For the construction of a DNA fusion between the *choB* gene coding for cholesterol oxidase (EC 1.1.3.6) and the 3' part of the AG- $\alpha$ 1 gene, the PCR technique on chromosomal DNA can be applied. Chromosomal DNA can be isolated by standard techniques from *Brevibacterium sterolicum*, and the DNA part coding for the mature part of the cholesterol oxidase can be amplified through application with the following corresponding PCR primers cho01pcr and cho02pcr:

cho01pct

5'-GCC CCC AGC CGC ACC CTC G-3' = SEQ ID NO: 16  
3'-CGG GGG TCG GCG TGG GAG C-5' = SEQ ID NO: 17  
5'-AGATCTGAATTCGCGGCC-GCC CCC AGC CGC ACC CTC G-3' = SEQ ID NO: 18  
EcoRI NotI  
EcoI

cho02pcr

NheI   HindIII  
 3'-TAG TAG AGC AGG CTG TAG GTC CGATCGACTTTCGAATCTAGA-5'      = SEQ ID NO: 19  
       |||    |||    |||    |||    |||    |||    |||  
 5'-ATC ATC TCG TCC GAC ATC CAG-3'      = SEQ ID NO: 20  
 3'-TAG TAG AGC AGG CTG TAG GTC-5'      = SEQ ID NO: 21

**[0065]** Both primers can specifically hybridize with the target sequence, thereby amplifying the coding part of the gene in such a way, that the specific PCR product -after Proteinase K treatment and digestion with *EcoRI* and *HindIII*- can be directly cloned into a suitable vector, here preferably pTZ19R, see Mead c.s. (1986). This will result in plasmid pUR2985 (see Figure 13).

**[0066]** In addition to the already mentioned restriction sites both PCR primers generate other restriction sites at the 5' end and the 3' end of the 1.5 kbp DNA fragment, which can be used later on to fuse the fragment in frame between either the SUC2 signal sequence or the prepro- $\alpha$ -mating factor signal sequence on one side and the C-terminus coding part of the  $\alpha$ -agglutinin gene on the other side. To facilitate the ligation behind the prepro-MF sequence a *NotI* site is introduced at the 5' end of PCR oligonucleotide cho1pcr, allowing for example, the exchange of the 731 bp *EagI/NheI* fragment containing the scFv-Lys coding sequence in pUR4175 for the *choB* coding sequence.

**[0067]** To create an enzymatically inactive fusion protein between cholesterol oxidase and  $\alpha$ -agglutinin, the above described subcloning into pTZ19R can be used. Cholesterol oxidase is an FAD-dependent enzyme for which the crystal structure of the *Brevibacterium sterolicum* enzyme has been determined, see [Vrielink c.s. \(1991\)](#). The enzyme displays homology with the typical pattern of the FAD-binding domain with the Gly-X-Gly-X-X-Gly sequence near the N-terminus (amino acid 18-23). Site-directed *in vitro* mutagenesis on the plasmid pUR2985 according to the manufacturer's protocol (Muta-Gene kit, Bio-Rad) can be applied to inactivate the FAD-binding site through replacing the triplet(s) encoding the Gly residue(s) by triplets encoding other amino acids, thereby presumably inactivating the enzyme. E.g. the following primer can be used for site-directed mutagenesis of 2 of the conserved Gly residues.

```

pr 3'- CGG GAG CAG TAG CGG TCA CGT ATG CCG CCA CGG CAG CCG CGC -5'
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
cs 5'- GCC CTC GTC ATC GGC AGT GGA TAC GGC GGT GCC GTC GCC GGC -3'
      Ala      Gly      Gly      Gly Gly Ala      Ala Ala
              ↓      ↓
              Ala      Ala

```

pr = primer = SEQ ID NO: 22  
cs = coding strand = SEQ ID NO: 23

**[0068]** As a result of the mutagenesis with the described primer, plasmid pUR2986 will be obtained. From this plasmid the DNA coding for the presumably inactivated cholesterol oxidase can be released as a 1527 bp fragment through *NotI/NheI* digestion, and subsequently directly used to exchange the scFv-Lys coding sequence in pUR4175, thereby generating plasmid pUR2987 (see Figure 14). To obtain a variant yeast secretion vector, where the secretion is directed through the SUC2 signal sequence, for example the 1823 bp long *SacI/NheI* segment of plasmid pUR2986 can be used to replace the *SacI/NheI* fragment in pUR4174.

**[0069]** This inactivation of the FAD-binding site might be preferable over other mutations, since an unchanged active centre can be expected to leave the binding properties of cholesterol oxidase for cholesterol unaltered. Instead of the described Gly-Ala exchanges at position 18 and 20 of the mature coding sequence, every other suitable amino acid change can also be performed.

**[0070]** To inactivate the enzyme, site directed mutagenesis can be optionally immediately performed in the active

**[0071]** It has been described that proteinase of *Lactococcus lactis* subsp. *cremoris* is anchored to the cell wall through its 127 amino acid long C-terminal, see Kok c.s. (1988) and Kok (1990). In a way similar to that described in Example 6, the cholesterol oxidase of *Brevibacterium sterolicum* (*choB*) can be immobilized on the surface of *Lactococcus lactis*. Fusions can be made between the *choB* structural gene and the N-terminal signal sequence and the C-terminal anchor of the proteinase of *Lactococcus lactis*. Plasmid pGKV550 (see Figure 15) contains the complete proteinase operon of *Lactococcus lactis* subsp. *cremoris* Wg2, including the promoter, a ribosome binding site and DNA fragments encoding the already mentioned signal and anchor sequences, see Kok (1990). First a DNA fragment, containing the main part of the signal sequence, flanked by a *Clal* site and an *EagI* site can be constructed with PCR on pGKV550 as follows:

3'-TT CCC GAT AGC TAG AAC AAT CGG CCA TGT CAG-5'  
                    ClaI  
= SEQ ID NO: 25

3'-CAG CCG CTT TAG GTT CGT TGC CGG CCC CCC TTC GAA CCC-5'  
EaqI HindIII

**[0073]** A copy of the cholesterol oxidase gene, suitable for fusion with the prtP anchor domain can be produced by a PCR reaction using plasmid pUR2985 (Example 6) as template and a combination of primer cho01pcr (see Example 6) and the following primer cho03pcr instead of primer cho02pcr:

= SEQ ID NO: 20.

**[0075]** After introduction into *Lactobacillus lactis* subsp. *lactis* MG1363 by electroporation, this plasmid will express cholesterol oxidase under control of the proteinase promoter. The transport through the membrane will be mediated by the proteinase signal sequence and the immobilization of the cholesterol oxidase by the proteinase anchor. As it is unlikely that the *Lactococcus* will secrete FAD as well, the cholesterol oxidase will not be active but will be capable to

bind cholesterol.

**Example 8. Construction of a gene encoding a chimeric protein that will be anchored in the cell wall of a lower eukaryote and is able to bind growth hormones, such as the epidermal growth factor.**

[0076] For the isolation of larger amounts of human epidermal growth factor (EGF) the corresponding receptor can be used in form of a fusion between the binding domain and a C-terminal part of  $\alpha$ -agglutinin as cell wall anchor. The complete cDNA sequence of the human epidermal growth factor is cloned and sequenced. For the construction of a fusion protein with EGF binding capacity the N-terminal part of the mature receptor until the central 23 amino acids, transmembrane region can be utilized.

[0077] The plasmid pUR4175 can be used for the construction. Through digestion with *EagI* and *NheI* (partial) a 731 bp DNA fragment containing the sequence coding for scFv is released and can be replaced by a DNA fragment coding for the first 621 amino acids of human epidermal growth factor receptor. Initiating from an existing human cDNA library or otherwise through production of a cDNA library by standard techniques from preferentially EGF receptor overexpressing cells, e.g. A431 carcinoma cells, see Ullrich c.s. (1984), further PCR can be applied for the generation of in frame linkage between the extracellular binding domain of the human growth factor receptor (amino acid 1-622) and the C-terminal part of  $\alpha$ -agglutinin.

[0078] PCR oligonucleotides for the in frame linkage of human epidermal growth factor receptor and the C-terminus of  $\alpha$ -agglutinin.

a: PCR oligonucleotides for the transition between SUC2 signal sequence and the N-terminus of mature EGF receptor.

```

                > mature EGF receptor
pri EGF1:      Ala Leu Glu   Lys Lys Val   = SEQ ID NO: 28
5'-GGG GCG GCC GCG CTG GAG GAA AAG AAA GTT TGC-3'
      NotI
3'-CGC TCA GCC CGA GAC CTC CTT TTC TTT CAA ACG 5'
EGF rec (non-coding strand): = SEQ ID NO: 29
  
```

b: PCR oligonucleotides for the in frame transition between C terminus of the extracellular binding domain of EGF receptor and the C terminal part of  $\alpha$ -agglutinin.

```

EGF rec (coding strand):
      Asn Gly Pro   Ile Pro Ser   Ala Thr
5'-AAT GGG CCT AAG ATC CCG TCC ATC GCC ACT-3' = SEQ ID NO: 30
      ||| ||| ||| ||| ||| ||| ||| = SEQ ID NO: 31
3'-TTA CCC GGA TTC TAG GGC AGG CGA TCGGAATTCGAA CCCC-5'
pr EGF2:      NheI HindIII
  
```

[0079] This fusion would result in an addition of 2 Ala amino acids between the signal sequence and the mature N-terminus of EGF receptor.

[0080] The newly obtained 1.9 kbp PCR fragment can be digested with *NotI* and *NheI* and directly ligated into the vector pUR4175 after digesting with the same enzymes, resulting in plasmid pUR2993 (see Figure 17), comprising the GAL7 promoter, the prepro- $\alpha$ -mating factor sequence, the chimeric EGF receptor binding domain gene /  $\alpha$ -agglutinin gene, the yeast 2  $\mu$ m sequence, the defective LEU2 promoter and the LEU2 gene. This plasmid can be transformed into *S. cerevisiae* and the transformed cells can be cultivated in YP medium whereby expression of the chimeric protein can be induced by adding galactose to the medium.

**Example 9. Construction of genes encoding a chimeric protein anchored to the cell wall of yeast, comprising a binding domain of a "Camelidae" heavy chain antibody**

[0081] Recently it was described that camels as well as a number of related species (e.g. llamas) contain a considerable amount of IgG antibody molecules which are only composed of heavy-chain dimers, see Hamers-Casterman

c.s. (1993). Although these "heavy-chain" antibodies are devoid of light chains, it was demonstrated, that they nevertheless have an extensive antigen-binding repertoire. In order to show that the variable regions of this type of antibodies can be produced and will be linked to the exterior of the cell wall of a yeast, the following constructs were prepared.

#### 5 Construction of pUR2997, pUR2998 and pUR2999

[0082] The about 2.1 kbp *EagI*-*HindIII* fragment of pUR4177 (Example 4, Fig 9) was isolated. By using PCR technology, an *EcoRI* restriction site was introduced immediately upstream of the *EagI* site, whereby the C of the *EcoRI* site is the same as the first C of the *EagI* site. The thus obtained *EcoRI*-*HindIII* fragment was ligated into plasmid

10 pEMBL9, which was digested with *EcoRI* and *HindIII*, which resulted in pUR4177.A

[0083] The *EcoRI*/*NheI* fragment of plasmid pUR4177.A was replaced by the *EcoRI*/*NheI* fragments of three different synthetic DNA fragments (SEQ ID NO: 32, SEQ ID NO: 33, and SEQ ID NO: 34) resulting in pUR2997, pUR2998 and pUR2999, respectively. The about 1.5 kbp *BstEII*-*HindIII* fragments of pUR2997 and pUR2998 were isolated.

#### 15 Construction of pUR4421

[0084] The multiple cloning site of plasmid pEMBL9, see Dente c.s. (1983), (ranging from the *EcoRI* to the *HindIII* site) was replaced by a synthetic DNA fragment having the nucleotide sequence given below, see SEQ ID NO: 35 giving the coding strand and SEQ ID NO: 36 giving the non-coding strand. The 5'-part of this nucleotide sequence comprises an *EagI* site, the first 4 codons of a *Camelidae* V<sub>H</sub> gene fragment (nucleotides 16-27) and a *XhoI* site (CTCGAG) coinciding with codons 5 and 6 (nucleotides 28-33). The 3'-part comprises the last 5 codons of the *Camelidae* V<sub>H</sub> gene (nucleotides 46-60) (part of which coincides with a *BstEII* site), eleven codons of the Myc tail (nucleotides 61-93), see SEQ ID NO: 35 containing these eleven codons and SEQ ID NO: 37 giving the amino acid sequence, and an *EcoRI* site (GAATTC). The *EcoRI* site, originally present in pEMBL9, is not functional any more, because the 5'-end of the nucleotide sequence contains AATTT instead of AATTC, indicated below as (*EcoRI*). The resulting plasmid is called pUR4421. The *Camelidae* V<sub>H</sub> fragment starts with amino acids Q-V-K and ends with amino acids V-S-S.

	( <i>EcoRI</i> )	<i>EagI</i>		<i>XhoI</i>		<i>BstEII</i>	
30	5'-	AATTTAGCGG	CGCCCCAGGT	GAAACTGCTC	GAGTAAGTGA	CTAAGGTCAC-	50
	3' 1	ATCGCC	GGCGGGTCCA	CTTTGACGAG	CTCATTCACT	GATTCCAGTG-	
		5		Q V K			
		-CGTCTCCTCA	GAACAAAAC	TCATCTCAGA	AGAGGATCTG	AATTAATGAG-	100
35		-GCAGAGGAGT	CTTGTTTTTG	AGTAGAGTCT	TCTCCTAGAC	TTAATTACTC-	
		V S S	E Q K	L I S E	E D L	N * *	
						= SEQ ID NO: 37	
	<i>EcoRI</i>		<i>HindIII</i>				
	-AATTCATCAA	ACGGTGATA		-3'	119	= SEQ ID NO: 35	
40	-TTAAGTAGTT	TGCCACTATT	CGA	-5'	123	= SEQ ID NO: 36	

#### Construction of pUR4424

45 [0085] After digesting the plasmid pB09 with *XhoI* and *BstEII*, a DNA fragment of about 0.34 kbp was isolated from agarose gel. This fragment codes for a truncated V<sub>H</sub> fragment, missing both the first 4 and the last 5 amino acids of the *Camelidae* V<sub>H</sub> fragment. Plasmid pB09 was deposited as *E. coli* JM109 pB09 at the Centraal Bureau voor Schimmelcultures, Baarn on 20 April 1993 with deposition number CBS 271.93. The DNA and amino acid sequences of the Camel V<sub>H</sub> fragments followed by the Flag sequence as present in plasmid pB09 were given in Figure 6B of European

50 patent application 93201239.6 (not yet published), which is herein incorporated by reference. The obtained about 0.34 kbp fragment was cloned into pUR4421. To this end plasmid pUR4421 was digested with *XhoI* and *HindIII*, after which the about 4 kb vector fragment was isolated from an agarose gel. The resulting vector was ligated with the about 0.34 kbp *XhoI*/*BstEII* fragment and a synthetic DNA linker having the following sequence:

55	<i>BstEII</i>		<i>HindIII</i>	
	GTCACCGTCTCCTCATAATGA			= SEQ ID NO: 38
	GCAGAGGAGTATTACTTCA			= SEQ ID NO: 39

resulting in plasmid pUR4421-09.

[0086] Plasmid pSY16 was digested with *EagI* and *HindIII*, after which the about 6.5 kbp long vector backbone was isolated and ligated with the about 0.38 kbp *EagI/HindIII* fragment from pUR4421-09 resulting in pUR4424.

#### 5 Construction of pUR4482 and pUR4483

[0087] From pUR4424 the about 0.44 kbp *SacI-BstEII* fragment, coding for the invertase signal sequence and the camel heavy chain variable 09 (= CH<sub>V</sub>09) fragment, was isolated as well as the about 6.3 kbp *SacI-HindIII* vector fragment. The about 6.3 kbp fragment and the about 0.44 kbp fragment from pUR4424 were ligated with the *BstEII-HindIII* fragment from pUR2997 or pUR2998 yielding pUR4482 and pUR4483, respectively.

[0088] Plasmid pUR4482 is thus an yeast episomal expression plasmid for expression of a fusion protein with the invertase signal sequence, the CH<sub>V</sub>09 variable region, the Myc-tail and the Camel "X-P-X-P" Hinge region, see Hamers-Casterman c.s. (1993), (1993), and the  $\alpha$ -agglutinin cell wall anchor region. Plasmid pUR4483 differs from pUR4482 in that it contains the Myc-tail but not the "X-P-X-P" Hinge region. Similarly, the *BstEII-HindIII* fragment from pUR2999 can be ligated with the about 6.3 kbp vector fragment and the about 0.44 kbp fragment from pUR4424, resulting in pUR4497, which will differ from pUR4482 in that it contains the "X-P-X-P" Hinge region but not the Myc-tail.

[0089] The plasmids pUR4424, pUR4482 and pUR4483 were introduced into *Saccharomyces cerevisiae* SU10 by electroporation, and transformants were selected on plates lacking leucine. Transformants from SU10 with pUR4424, pUR4482 or pUR4483, respectively, were grown on YP with 5% galactose and analysed with immune-fluorescence microscopy, as described in Example 1 of our co-pending WO-94/01567 (UNILEVER) published on 20 January 1994. This method was slightly modified to detect the chimeric proteins, containing both the camel antibody and the Myc tail, present at the cell surface.

[0090] In one method a monoclonal mouse anti-Myc antibody was used as a first antibody to bind to the Myc part of the chimeric protein; subsequently a polyclonal antmouse 1g antiserum labeled with fluorescein isothiocyanate (= FITC) ex Sigma, Product No. F-0527, was used to detect the bound mouse antibody and a positive signal was determined by fluorescence microscopy.

[0091] In the other method a polyclonal rabbit anti-human IgG serum, which had earlier been proven to cross-react with the camel antibodies, was used as a first antibody to bind the camel antibody part of the chimeric protein; subsequently a polyclonal anti-rabbit 1g antiserum labeled with FITC ex Sigma, Product No. F-0382, was used to detect the bound rabbit antibody and a positive signal was determined by fluorescence microscopy.

[0092] The results in Figure 19 and Figure 20 show clearly that fluorescence can be observed on those cells in which a fusion protein of the CH<sub>V</sub>09 fragment with the  $\alpha$ -agglutinin cell wall anchor region is produced (pUR4482 and pUR4483). No fluorescence however, was visible on the cells which produce the CH<sub>V</sub>09 fragment without this anchor (pUR4424), when viewed under the same circumstances.

#### Patent literature references:

[0093] EP-A1-0255153 (UNILEVER) Production of guar alpha-galactosidase by hosts transformed by recombinant DNA methods. First priority date 03.06.86; published 03.02.88

[0094] WO-91/00920 (UNILEVER) Process for preparing a protein by a fungus transformed by multicopy integration of an expression vector. First priority date 07.07.89; published 24.01.91

[0095] WO-91/19782 (UNILEVER) Xylanase production. Priority date 19.06.90; published 26.12.91

[0096] WO-94/01567 (UNILEVER) Process for immobilizing enzymes to the cell wall of a microbial cell by producing a fusion protein. First priority date 08.07.92; published 20.01.94

[0097] EP patent application 93201239.6 (not yet published) Production of antibodies or (functionalized) fragments thereof derived from heavy chain immunoglobulins of *Camelidae*. Filing date 29.04.93

#### Non-patent literature references:

[0098] R.E. Bird & B. Webb Walker Single chain antibody variable regions. *TIBTECH* 9 (April 1991) 132-137

A. Conzelmann, C. Fankhauser & C. Desponds Myoinositol gets incorporated into numerous membrane glycoproteins of *Saccharomyces cerevisiae*; incorporation is dependent on phosphomannomutase (SEC53). *The EMBO Journal* 9, No.3 (1990) 653-661

L. Dente, G. Cesareni & R. Cortese pEMBL: a new family of single stranded plasmids. *Nucleic Acids Research* 11, No. 6 (1983) 1645-1655

E. Erhart, & C.P. Hollenberg The Presence of a Defective *Leu2* Gene on 2 $\mu$  DNA Recombinant Plasmids of *Saccharomyces cerevisiae* Is Responsible for Curing and High Copy Number. *Journal of Bacteriology*. 156, No.2 (Nov. 1983) 625-635



C. Hamers-Casterman, T. Atarhouch, S. Muyldermans, G. Robinson, C. Hamers, E. Bajyana Songa, N. Bendahman & R. Hamers Naturally occurring antibodies devoid of light chains. *Nature* **363** (3 June 1993) 446-448.

M.M. Harmsen, A.C. Langedijk, E. van Tuinen, R.H. Geerse, H.A. Raué, & J. Maat Effect of pmrl disruption and different signal sequences on the intracellular processing and secretion of *Cyamopsis tetragonoloba*  $\alpha$ -galactosidase by *Saccharomyces cerevisiae*. *Gene* **125** (1993) 115-123

T. Ishizaki, N. Hirayama, H. Shinkawa, O. Nimi, & Y. Murooka Nucleotide Sequence of the Gene for Cholesterol Oxidase from a *Streptomyces* sp. *Journal of Bacteriology* **171**, No.1 (Jan. 1989) 596-601

J. Kok, K.J. Leenhouts, A.J. Haandrikman, A.M. Ledeboer, & G. Venema Nucleotide Sequence of the Cell Wall Proteinase Gene of *Streptococcus cremoris* Wg2. *Applied and Environmental Microbiology* **54**, No. 1 (Jan. 1988) 231-238

J. Kok Genetics of the proteolytic system of lactic acid bacteria. *FEMS Microbiology Reviews*. **87** (1990) 15-42, esp. item 3.2 and Figure 7 on pages 22-23 R. Lewin The universal constructor set. *New Scientist* (8 December 1990) 30-33

P.N. Lipke, D. Wojciechowicz & J. Kurjan AGA1 Is the Structural Gene for the *Saccharomyces cerevisiae*  $\alpha$ -Agglutinin, a Cell Surface Glycoprotein Involved in Cell-Cell Interactions during Mating. *Molecular and Cellular Biology*. **9**, No.8 (Aug. 1989) 3155-3165.

D.A. Mead, E. Szczesna-Skorupa & B. Kemper Single-stranded DNA 'blue' T7 promoter plasmids: a versatile tandem promoter system for cloning and protein engineering. *Protein Engineering* **1**, No. 1 (1986) 67-74

R. Orlandi, D.H. Güssow, P.T. Jones, & G. Winter Cloning immunoglobulin variable domains for expression by the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **86** (May 1989) 3833-3837

B.C. Ossendorp The non-specific lipid-transfer binding protein: a recombinant DNA and immunological approach. Thesis University of Utrecht (30 September 1992) T. Ohta, K. Fujishiro, K. Yamaguchi, Y. Tamura, K. Aisaka, T. Uwajima & M. Hasegawa Sequence of gene *choB* encoding cholesterol oxidase of *Brevibacterium sterolicum*: comparison with *choA* of *Streptomyces* sp. SA-COO. *Gene* **103** (1991) 93-96

A. Roy, C.F. Lu, D.L. Marykwas, P.N. Lipke & J. Kurjan The AGA1 Product Is Involved in Cell Surface Attachment of the *Saccharomyces cerevisiae* Cell Adhesion Glycoprotein  $\alpha$ -Agglutinin. *Molecular and Cellular Biology* **11** No.8 (Aug. 1991) 4196-4206

M.P. Schreuder, S. Brekelmans, H. van den Ende, & F.M. Klis Targeting of a Heterologous Protein to the Cell Wall of *Saccharomyces cerevisiae*. *Yeast* **9** (1993) 399-409

A. Ullrich, L. Coussens, J.S. Hayflick, T.J. Dull, A. Gray, A.W. Tam, J. Lee, Y.

Yarden, T.A. Libermann, J. Schlessinger, J. Downward, E.L.V. Mayes, N. Whittle,

M.D. Waterfield & P.H. Seeburg Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. *Nature* **309** (31 May 1984) 418-425

J.M.A. Verbakel Heterologous gene expression in the yeast *Saccharomyces cerevisiae*. Thesis University of Utrecht (1 May 1991), esp. pages 76-89

A. Vrielink, L.F. Lloyd, & D.M. Blow Crystal Structure of Cholesterol Oxidase from *Brevibacterium sterolicum* Refined at 1.8 Å Resolution. *J. Mol. Biol.*, **219** (1991) 533-554

E.S. Ward, D. Güssow, A.D. Griffiths, P.T. Jones & G. Winter. Binding activities of a repertoire of single immunoglobulin variable domains secreted from *E. coli*. *Nature* **341** (12 October 1989) 544-546

[0099] Information on a deposit of a micro-organism under the Budapest Treaty is given on page 26, lines 5-7 above. In agreement with Rule 28 (4) EPC, or a similar arrangement for a State not being a Contracting State of the EPC, it is hereby requested that a sample of such deposit, when requested, will be submitted to an expert only.

## SEQUENCE LISTING

[0100]

### (1) GENERAL INFORMATION:

#### (i) APPLICANT:

- (A) NAME: Unilever N.V.
- (B) STREET: Weena 455
- (C) CITY: Rotterdam
- (E) COUNTRY: The Netherlands
- (F) POSTAL CODE (ZIP): NL-3013 AL

- (A) NAME: Unilever PLC
- (B) STREET: Unilever House Blackfriars

(C) CITY: London  
(E) COUNTRY: United Kingdom  
(F) POSTAL CODE (ZIP): EC4P 4BQ

5 (A) NAME: Leon Gerardus J. FRENKEN  
(B) STREET: Geldersestraat 90  
(C) CITY: Rotterdam  
(E) COUNTRY: The Netherlands  
10 (F) POSTAL CODE (ZIP): NL-3011 MP

(A) NAME: Pieter DE GEUS  
(B) STREET: Boeier 24  
(C) CITY: Barendrecht  
(E) COUNTRY: The Netherlands  
15 (F) POSTAL CODE (ZIP): NL-2991 KB

(A) NAME: Franciscus Maria KLIS  
(B) STREET: Benedenlangs 102  
(C) CITY: Amsterdam  
20 (E) COUNTRY: The Netherlands  
(F) POSTAL CODE (ZIP): NL-1025 KL

(A) NAME: Holger York TOSCHKA; c/o Langnese Iglo, BR3  
(B) STREET: Aeckern 1  
25 (C) CITY: REKEN  
(E) COUNTRY: Germany  
(F) POSTAL CODE (ZIP): D-48734

(A) NAME: cornelis Theodorus VERRIPS  
30 (B) STREET: Hagedoorn 18  
(C) CITY: Maassluis  
(E) COUNTRY: The Netherlands  
(F) POSTAL CODE (ZIP): NL-3142 KB

35 (ii) TITLE OF INVENTION: Immobilized proteins with specific binding capacities and their use in processes and products.

(iii) NUMBER OF SEQUENCES: 40

40 (iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
45 (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

50 (A) LENGTH: 231 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

(B) CLONE: fragment in pUR4119

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

5  
GAATTCGAGC TCATCACACA AACAAACAAA ACAAAATGAT GCTTTTGCAA GCCTTTCTTT 60  
TCCTTTTGGC TGGTTTTGCA GCCAAAATAT CTGCGCAGGT GCAGCTGCAG TAATGAACCA 120  
10 CCGTCACCGT CTCCTCAGGT GGAGGCGGTT CAGGCGGAGG TGGCTCTGGC GGTGGCGGAT 180  
CGGACATCGA GCTCACTCAG ACCAAGCTCG AGATCAAACG GTGATAAGCT T 231

(2) INFORMATION FOR SEQ ID NO: 2:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

20

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25

(vii) IMMEDIATE SOURCE:

(B) CLONE: linker XhoI-NheI coding strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

30

TOGAGATCAA AGGCGGATCT G 21

(2) INFORMATION FOR SEQ ID NO: 3:

35

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

40

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

45

(vii) IMMEDIATE SOURCE:

(B) CLONE: linker XhoI-NheI non-coding strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

50

CTAGCAGATC CGCCTTTGAT C 21

(2) INFORMATION FOR SEQ ID NO: 4:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

5 (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

(B) CLONE: linker EagI-PstI coding strand

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

**GGCCGCCCCAG GTGCAGCTGC A**

**21**

15

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

20

(A) LENGTH: 13 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

30

(B) CLONE: linker EagI-PstI non-coding strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

35

**GCTGCACCTG GGC**

**13**

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

40

(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

45

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

50

(B) CLONE: PCR primer A (heavy chain)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

55

**AGGTSMARCT GCAGSAGTCW GG**

**22**

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: PCR primer B (heavy chain)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

**TGAGGAGACG GTGACCGTGG TCCCTTGGCC CC**

**32**

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: PCR primer C (light chain)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

**GACATTGAGC TCACCCAGTC TCCA**

**24**

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: PCR primer D (light chain)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

**GTTTGATCTC GAGCTTGGTC CC**

**22**

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 26 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

10 (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: linker EcoRI-PatI coding strand

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

**AATTCGGCCG TTCAGGTGCA CCTGCA**

**26**

20

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- 25 (A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

30 (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: linker EcoRI-PstI non-coding strand

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

**GCTGCACCTG AACGGCCG**

**18**

40

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 714 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50 (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- (B) CLONE: ScFv antitransolide 02/01/01

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

	CTGCAGGAGT CTGGACCTGG CCTGGTGAAA CCTTCTCAGT CTCTGTCCCT CACCTGCACT	60
	GTCACCTGGCT ACTCAATCAC CAGTGATTTT GCCTGGAAGT GGATCCGGCA GTTTCAGGA	120
5	AACCAACTGG AGTGGATGGG CTACATAAGC TACAGTGGTA GCACTAGCTA CAACCCATCT	180
	CTCAAAAGTC GAATCTCTCT CACTCGAGAC ACATCCAAGA ACCAGTTCTT CCTGCAGTTG	240
	AATTCTGTGA CTACTGAGGA CACAGCCACA TATTACTGTG CAACGTCCCT AACATGGTTA	300
10	CTACGTGGGA AACGTTCTTA CTGGGGCCAA GGGACCACGG TCACCGTCTC CTCAGGTGGA	360
	GGCGGTTTCAG GCGGAGGTGG CTCTGGCGGT GCGGGATCGG ACATCGAGCT CACCCAGTCT	420
	CCATCCTCCA TGTCTGTATC TCTGGGAGAC ACAGTCAGCA TCACTTGCCA TGCAAGTCAG	480
15	GACATTAGCA GTAATATAGG GTGGTTGCAG CAGAAACCAG GGAAATCATT TAAGGGCCTG	540
	ATCTATCATG GAACCAACTT GGAAGATGCT ATTCCATCAA GGTTCAGTGG CAGTGGATCT	600
	GGAGCAGATT ATTCCCTCAC CATCAGCAGC CTGGAATCTG AACATTTTGC AGACTATTAC	660
20	TGTGTACAGT ATGCTCAGTT TCCATTCACG TTCGGCTGGG GGACCAAGCT CGAG	714

## (2) INFORMATION FOR SEQ ID NO: 13:

## 25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 734 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vii) IMMEDIATE SOURCE:

- (B) CLONE: ScFv anti-HCG

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

40

45

50

55

CGGCCGTTCA GGTGCAGCTG CAGGAGTCTG GGGGACACTT AGTGAAGCCT GGAGGGTCCC 60  
 TGAAGTCTC CTGTGCAGCC TCTGGATTGG CTTTCAGTAG CTTTGACATG TCTTGGATTG 120  
 5 GCCAGACTCC GGAGAAGAGG CTGGAGTGGG TCGCAAGCAT TACTAATGTT GGTACTTACA 180  
 CCTACTATCC AGGCAGTGTG AAGGGCCGAT TCTCCATCTC CAGAGACAAT GCCAGGAACA 240  
 CCCTAAACCT GCAAATGAGC AGTCTGAGGT CTGAGGACAC GGCCTTGTAT TTCTGTGCAA 300  
 10 GACAGGGGAC TCGGGCACAA CCTTACTGGT ACTTCGATGT CTGGGGCCAA GGGACCAOCC 360  
 TCACCGTCTC CTCAGGTGGA GCGGTTTCTG GCGGAGGTGG CTCTGGCGGT GCGGATCCG 420  
 ACATCCAGCT CACCCAGTCT CCAAATCCA TGTCCATGTC CGTAGGAGAG AGGGTCACCT 480  
 15 TGAGCTGCAA GGCCAGTGAG ACTGTGGATT CTTTTGTGTC CTGGTATCAA CAGAAACCG 540  
 AACAGTCTCC TAAATTGTTG ATATTGGGG CATCCAACCG GTTCAGTGGG GTCCCCGATC 600  
 GCTTCACTGG CAGTGGATCT GCAACAGACT TCACTCTGAC CATCAGCAGT GTGCAGGCTG 660  
 20 AGGACTTTGC GGATTACCAC TGTGGACAGA CTTACAATCA TCCGTATACG TTCGGAGGGG 720  
 GGACCAAGCT CGAG 734

25

(2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 2685 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

40

- (A) ORGANISM: *Saccharomyces cerevisiae*

## (vii) IMMEDIATE SOURCE:

45

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..2685
- (D) OTHER INFORMATION: /product= "Flocculation protein" /gene= "FLO1"

50

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

55



	ATG ACA ATG CCT CAT CGC TAT ATG TTT TTG GCA GTC TTT ACA CTT CTG	48
	Met Thr Met Pro His Arg Tyr Met Phe Leu Ala Val Phe Thr Leu Leu	
	1 5 10 15	
5	GCA CTA ACT AGT GTG GCC TCA GGA GCC ACA GAG GCG TGC TTA CCA GCA	96
	Ala Leu Thr Ser Val Ala Ser Gly Ala Thr Glu Ala Cys Leu Pro Ala	
	20 25 30	
10	GGC CAG AGG AAA AGT GGG ATG AAT ATA AAT TTT TAC CAG TAT TCA TTG	144
	Gly Gln Arg Lys Ser Gly Met Asn Ile Asn Phe Tyr Gln Tyr Ser Leu	
	35 40 45	
	AAA GAT TCC TCC ACA TAT TCG AAT GCA GCA TAT ATG GCT TAT GGA TAT	192
	Lys Asp Ser Ser Thr Tyr Ser Asn Ala Ala Tyr Met Ala Tyr Gly Tyr	
	50 55 60	
15	GCC TCA AAA ACC AAA CTA GGT TCT GTC GGA GGA CAA ACT GAT ATC TCG	240
	Ala Ser Lys Thr Lys Leu Gly Ser Val Gly Gly Gln Thr Asp Ile Ser	
	65 70 75 80	
20	ATT GAT TAT AAT ATT CCC TGT GTT AGT TCA TCA GGC ACA TTT CCT TGT	288
	Ile Asp Tyr Asn Ile Pro Cys Val Ser Ser Gly Thr Phe Pro Cys	
	85 90 95	
	CCT CAA GAA GAT TCC TAT GGA AAC TGG GGA TGC AAA GGA ATG GGT GCT	336
	Pro Gln Glu Asp Ser Tyr Gly Asn Trp Gly Cys Lys Gly Met Gly Ala	
	100 105 110	
25	TGT TCT AAT AGT CAA GGA ATT GCA TAC TGG AGT ACT GAT TTA TTT GGT	384
	Cys Ser Asn Ser Gln Gly Ile Ala Tyr Trp Ser Thr Asp Leu Phe Gly	
	115 120 125	
30	TTC TAT ACT ACC CCA ACA AAC GTA ACC CTA GAA ATG ACA GGT TAT TTT	432
	Phe Tyr Thr Thr Pro Thr Asn Val Thr Leu Glu Met Thr Gly Tyr Phe	
	130 135 140	
	TTA CCA CCA CAG ACG GGT TCT TAC ACA TTC AAG TTT GCT ACA GTT GAC	480
	Leu Pro Pro Gln Thr Gly Ser Tyr Thr Phe Lys Phe Ala Thr Val Asp	
	145 150 155 160	
35	GAC TCT GCA ATT CTA TCA GTA GGT GGT GCA ACC GCG TTC AAC TGT TGT	528
	Asp Ser Ala Ile Leu Ser Val Gly Gly Ala Thr Ala Phe Asn Cys Cys	
	165 170 175	

40

45

50

55

EP 0 682 710 B1

	GCT CAA CAG CAA CCG CCG ATC ACA TCA ACG AAC TTT ACC ATT GAC GGT	576
	Ala Gln Gln Gln Pro Pro Ile Thr Ser Thr Asn Phe Thr Ile Asp Gly	
	180 185 190	
5	ATC AAG CCA TGG GGT GGA AGT TTG CCA CCT AAT ATC GAA GGA ACC GTC	624
	Ile Lys Pro Trp Gly Gly Ser Leu Pro Pro Asn Ile Glu Gly Thr Val	
	195 200 205	
	TAT ATG TAC GCT GGC TAC TAT TAT CCA ATG AAG GTT GTT TAC TCG AAC	672
10	Tyr Met Tyr Ala Gly Tyr Tyr Tyr Pro Met Lys Val Val Tyr Ser Asn	
	210 215 220	
	GCT GTT TCT TGG GGT ACA CTT CCA ATT AGT GTG ACA CTT CCA GAT GGT	720
	Ala Val Ser Trp Gly Thr Leu Pro Ile Ser Val Thr Leu Pro Asp Gly	
	225 230 235 240	
15	ACC ACT GTA ACT GAT GAC TTC GAA GGG TAC GTC TAT TCC TTT GAC GAT	768
	Thr Thr Val Ser Asp Phe Glu Gly Tyr Val Tyr Ser Phe Asp Asp	
	245 250 255	
	GAC CTA AGT CAA TCT AAC TGT ACT GTC CCT GAC CCT TCA AAT TAT GCT	816
20	Asp Leu Ser Gln Ser Asn Cys Thr Val Pro Asp Pro Ser Asn Tyr Ala	
	260 265 270	
	GTC AGT ACC ACT ACA ACT ACA ACG GAA CCA TGG ACC GGT ACT TTC ACT	864
	Val Ser Thr Thr Thr Thr Thr Glu Pro Trp Thr Gly Thr Phe Thr	
	275 280 285	
25	TCT ACA TCT ACT GAA ATG ACC ACC GTC ACC GGT ACC AAC GGC GTT CCA	912
	Ser Thr Ser Thr Glu Met Thr Thr Val Thr Gly Thr Asn Gly Val Pro	
	290 295 300	
	ACT GAC GAA ACC GTC ATT GTC ATC AGA ACT CCA ACC AGT GAA GGT CTA	960
30	Thr Asp Glu Thr Val Ile Val Ile Arg Thr Pro Thr Ser Glu Gly Leu	
	305 310 315 320	
	ATC AGC ACC ACC ACT GAA CCA TGG ACT GGC ACT TTC ACT TCG ACT TCC	1008
	Ile Ser Thr Thr Thr Glu Pro Trp Thr Gly Thr Phe Thr Ser Thr Ser	
	325 330 335	
	ACT GAG GTT ACC ACC ATC ACT GGA ACC AAC GGT CAA CCA ACT GAC GAA	1056
35	Thr Glu Val Thr Thr Ile Thr Gly Thr Asn Gly Gln Pro Thr Asp Glu	
	340 345 350	
	ACT GTG ATT GTT ATC AGA ACT CCA ACC AGT GAA GGT CTA ATC AGC ACC	1104
	Thr Val Ile Val Ile Arg Thr Pro Thr Ser Glu Gly Leu Ile Ser Thr	
	355 360 365	
40	ACC ACT GAA CCA TGG ACT GGT ACT TTC ACT TCT ACA TCT ACT GAA ATG	1152
	Thr Thr Glu Pro Trp Thr Gly Thr Phe Thr Ser Thr Ser Thr Glu Met	
	370 375 380	
	ACC ACC GTC ACC GGT ACT AAC GGT CAA CCA ACT GAC GAA ACC GTG ATT	1200
45	Thr Thr Val Thr Gly Thr Asn Gly Gln Pro Thr Asp Glu Thr Val Ile	
	385 390 395 400	
	GTT ATC AGA ACT CCA ACC AGT GAA GGT TTG GTT ACA ACC ACC ACT GAA	1248
	Val Ile Arg Thr Pro Thr Ser Glu Gly Leu Val Thr Thr Thr Thr Glu	
	405 410 415	
50	CCA TGG ACT GGT ACT TTT ACT TCG ACT TCC ACT GAA ATG TCT ACT GTC	1296
	Pro Trp Thr Gly Thr Phe Thr Ser Thr Ser Thr Glu Met Ser Thr Val	
	420 425 430	
	ACT GGA ACC AAT GGC TTG CCA ACT GAT GAA ACT GTC ATT GTT GTC AAA	1344
55	Thr Gly Thr Asn Gly Leu Pro Thr Asp Glu Thr Val Ile Val Val Lys	
	435 440 445	

	ACT CCA ACT ACT GCC ATC TCA TCC AGT TTG TCA TCA TCA TCT TCA GGA	1392
	Thr Pro Thr Thr Ala Ile Ser Ser Ser Leu Ser Ser Ser Ser Ser Gly	
	450 455 460	
5	CAA ATC ACC AGC TCT ATC ACG TCT TCG CGT CCA ATT ATT ACC CCA TTC	1440
	Gln Ile Thr Ser Ser Ile Thr Ser Ser Arg Pro Ile Ile Thr Pro Phe	
	465 470 475 480	
10	TAT CCT AGC AAT GGA ACT TCT GTG ATT TCT TCC TCA GTA ATT TCT TCC	1488
	Tyr Pro Ser Asn Gly Thr Ser Val Ile Ser Ser Ser Val Ile Ser Ser	
	485 490 495	
	TCA GTC ACT TCT TCT CTA TTC ACT TCT TCT CCA GTC ATT TCT TCC TCA	1536
	Ser Val Thr Ser Ser Leu Phe Thr Ser Ser Pro Val Ile Ser Ser Ser	
	500 505 510	
15	GTC ATT TCT TCT TCT ACA ACA ACC TCC ACT TCT ATA TTT TCT GAA TCA	1584
	Val Ile Ser Ser Ser Thr Thr Ser Thr Ser Ile Phe Ser Glu Ser	
	515 520 525	
20	TCT AAA TCA TCC GTC ATT CCA ACC AGT AGT TCC ACC TCT GGT TCT TCT	1632
	Ser Lys Ser Ser Val Ile Pro Thr Ser Ser Ser Thr Ser Gly Ser Ser	
	530 535 540	
	GAG AGC GAA ACG AGT TCA GCT GGT TCT GTC TCT TCT TCC TCT TTT ATC	1680
	Glu Ser Glu Thr Ser Ser Ala Gly Ser Val Ser Ser Ser Ser Phe Ile	
	545 550 555 560	
25	TCT TCT GAA TCA TCA AAA TCT CCT ACA TAT TCT TCT TCA TCA TTA CCA	1728
	Ser Ser Glu Ser Ser Lys Ser Pro Thr Tyr Ser Ser Ser Ser Leu Pro	
	565 570 575	
30	GTT GTT ACC AGT GCG ACA ACA AGC CAG GAA ACT GCT TCT TCA TTA CCA	1776
	Leu Val Thr Ser Ala Thr Thr Ser Gln Glu Thr Ala Ser Ser Leu Pro	
	580 585 590	
	CCT GCT ACC ACT ACA AAA ACG AGC GAA CAA ACC ACT TTG GTT ACC GTG	1824
	Pro Ala Thr Thr Thr Lys Thr Ser Glu Gln Thr Thr Leu Val Thr Val	
	595 600 605	
35	ACA TCC TGC GAG TCT CAT GTG TGC ACT GAA TCC ATC TCC CCT GCG ATT	1872
	Thr Ser Cys Glu Ser His Val Cys Thr Glu Ser Ile Ser Pro Ala Ile	
	610 615 620	
	GTT TCC ACA GCT ACT GTT ACT GTT AGC GGC GTC ACA ACA GAG TAT ACC	1920
	Val Ser Thr Ala Thr Val Thr Val Ser Gly Val Thr Thr Glu Tyr Thr	
	625 630 635 640	
40	ACA TGG TGC CCT ATT TCT ACT ACA GAG ACA ACA AAG CAA ACC AAA GGG	1968
	Thr Trp Cys Pro Ile Ser Thr Thr Glu Thr Thr Lys Gln Thr Lys Gly	
	645 650 655	
45	ACA ACA GAG CAA ACC ACA GAA ACA ACA AAA CAA ACC ACG GTA GTT ACA	2016
	Thr Thr Glu Gln Thr Thr Glu Thr Thr Lys Gln Thr Thr Val Val Thr	
	660 665 670	
	ATT TCT TCT TGT GAA TCT GAC-GTA TGC TCT AAG ACT GCT TCT CCA GCC	2064
	Ile Ser Ser Cys Glu Ser Asp Val Cys Ser Lys Thr Ala Ser Pro Ala	
	675 680 685	
50	ATT GTA TCT ACA AGC ACT GCT ACT ATT AAC GGC GTT ACT ACA GAA TAC	2112
	Ile Val Ser Thr Ser Thr Ala Thr Ile Asn Gly Val Thr Thr Glu Tyr	
	690 695 700	
55	ACA ACA TGG TGT CCT ATT TCC ACC ACA GAA TCG ACG CAA CAA ACA ACG	2160
	Thr Thr Trp Cys Pro Ile Ser Thr Thr Glu Ser Arg Gln Gln Thr Thr	
	705 710 715 720	

	CTA GTT ACT GTT ACT TCC TGC GAA TCT GGT GTG TGT TCC GAA ACT GCT	2208
	Leu Val Thr Val Thr Ser Cys Glu Ser Gly Val Cys Ser Glu Thr Ala	
	725 730 735	
5	TCA CCT GCC ATT GTT TCG ACG GCC ACG GCT ACT GTG AAT GAT GTT GTT	2256
	Ser Pro Ala Ile Val Ser Thr Ala Thr Val Asn Asp Val Val	
	740 745 750	
10	ACG GTC TAT CCT ACA TGG AGG CCA CAG ACT GCG AAT GAA GAG TCT GTC	2304
	Thr Val Tyr Pro Thr Trp Arg Pro Gln Thr Ala Asn Glu Glu Ser Val	
	755 760 765	
	AGC TCT AAA ATG AAC AGT GCT ACC GGT GAG ACA ACA ACC AAT ACT TTA	2352
	Ser Ser Lys Met Asn Ser Ala Thr Gly Glu Thr Thr Thr Asn Thr Leu	
	770 775 780	
15	GCT GCT GAA ACG ACT ACC AAT ACT GTA GCT GCT GAG ACG ATT ACC AAT	2400
	Ala Ala Glu Thr Thr Thr Asn Thr Val Ala Ala Glu Thr Ile Thr Asn	
	785 790 795 800	
20	ACT GGA GCT GCT GAG ACG AAA ACA GTA GTC ACC TCT TCG CTT TCA AGA	2448
	Thr Gly Ala Ala Glu Thr Lys Thr Val Val Thr Ser Ser Leu Ser Arg	
	805 810 815	
	TCT AAT CAC GCT GAA ACA CAG ACG GCT TCC GCG ACC GAT GTG ATT GGT	2496
	Ser Asn His Ala Glu Thr Gln Thr Ala Ser Ala Thr Asp Val Ile Gly	
	820 825 830	
25	CAC AGC AGT AGT GTT GTT TCT GTA TCC GAA ACT GGC AAC ACC AAG AGT	2544
	His Ser Ser Ser Val Val Ser Val Ser Glu Thr Gly Asn Thr Lys Ser	
	835 840 845	
30	CTA ACA AGT TCC GGG TTG AGT ACT ATG TCG CAA CAG CCT CGT AGC ACA	2592
	Leu Thr Ser Ser Gly Leu Ser Thr Met Ser Gln Gln Pro Arg Ser Thr	
	850 855 860	
	CCA GCA AGC AGC ATG GTA GGA TAT AGT ACA GCT TCT TTA GAA ATT TCA	2640
	Pro Ala Ser Ser Met Val Gly Tyr Ser Thr Ala Ser Leu Glu Ile Ser	
	865 870 875 880	
35	ACG TAT GCT GGC AGT GCA ACA GCT TAC TGG CCG GTA GTG GTT TAA	2685
	Thr Tyr Ala Gly Ser Ala Thr Ala Tyr Trp Pro Val Val Val	
	885 890 895	

(2) INFORMATION FOR SEQ ID NO: 15:

40

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 894 amino acids

(B) TYPE: amino acid

45

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

50

55

EP 0 682 710 B1

	Met	Thr	Met	Pro	His	Arg	Tyr	Met	Phe	Leu	Ala	Val	Phe	Thr	Leu	Leu	
	1				5					10					15		
5	Ala	Leu	Thr	Ser	Val	Ala	Ser	Gly	Ala	Thr	Glu	Ala	Cys	Leu	Pro	Ala	
				20					25					30			
	Gly	Gln	Arg	Lys	Ser	Gly	Met	Asn	Ile	Asn	Phe	Tyr	Gln	Tyr	Ser	Leu	
			35					40					45				
10	Lys	Asp	Ser	Ser	Thr	Tyr	Ser	Asn	Ala	Ala	Tyr	Met	Ala	Tyr	Gly	Tyr	
		50					55					60					

15

20

25

30

35

40

45

50

55

Ala Ser Lys Thr Lys Leu Gly Ser Val Gly Gly Gln Thr Asp Ile Ser  
 65 70 75 80  
 5 Ile Asp Tyr Asn Ile Pro Cys Val Ser Ser Ser Gly Thr Phe Pro Cys  
 85 90 95  
 Pro Gln Glu Asp Ser Tyr Gly Asn Trp Gly Cys Lys Gly Met Gly Ala  
 100 105 110  
 10 Cys Ser Asn Ser Gln Gly Ile Ala Tyr Trp Ser Thr Asp Leu Phe Gly  
 115 120 125  
 Phe Tyr Thr Thr Pro Thr Asn Val Thr Leu Glu Met Thr Gly Tyr Phe  
 130 135 140  
 15 Leu Pro Pro Gln Thr Gly Ser Tyr Thr Phe Lys Phe Ala Thr Val Asp  
 145 150 155 160  
 Asp Ser Ala Ile Leu Ser Val Gly Gly Ala Thr Ala Phe Asn Cys Cys  
 165 170 175  
 20 Ala Gln Gln Gln Pro Pro Ile Thr Ser Thr Asn Phe Thr Ile Asp Gly  
 180 185 190  
 Ile Lys Pro Trp Gly Gly Ser Leu Pro Pro Asn Ile Glu Gly Thr Val  
 195 200 205  
 25 Tyr Met Tyr Ala Gly Tyr Tyr Tyr Pro Met Lys Val Val Tyr Ser Asn  
 210 215 220  
 Ala Val Ser Trp Gly Thr Leu Pro Ile Ser Val Thr Leu Pro Asp Gly  
 225 230 235 240  
 30 Thr Thr Val Ser Asp Asp Phe Glu Gly Tyr Val Tyr Ser Phe Asp Asp  
 245 250 255  
 Asp Leu Ser Gln Ser Asn Cys Thr Val Pro Asp Pro Ser Asn Tyr Ala  
 260 265 270  
 35 Val Ser Thr Thr Thr Thr Thr Thr Glu Pro Trp Thr Gly Thr Phe Thr  
 275 280 285  
 Ser Thr Ser Thr Glu Met Thr Thr Val Thr Gly Thr Asn Gly Val Pro  
 290 295 300  
 40 Thr Asp Glu Thr Val Ile Val Ile Arg Thr Pro Thr Ser Glu Gly Leu  
 305 310 315 320  
 Ile Ser Thr Thr Thr Glu Pro Trp Thr Gly Thr Phe Thr Ser Thr Ser  
 325 330 335  
 Thr Glu Val Thr Thr Ile Thr Gly Thr Asn Gly Gln Pro Thr Asp Glu  
 340 345 350  
 45 Thr Val Ile Val Ile Arg Thr Pro Thr Ser Glu Gly Leu Ile Ser Thr  
 355 360 365  
 Thr Thr Glu Pro Trp Thr Gly Thr Phe Thr Ser Thr Ser Thr Glu Met  
 370 375 380  
 50 Thr Thr Val Thr Gly Thr Asn Gly Gln Pro Thr Asp Glu Thr Val Ile  
 385 390 395 400  
 Val Ile Arg Thr Pro Thr Ser Glu Gly Leu Val Thr Thr Thr Thr Glu  
 405 410 415

EP 0 682 710 B1

Pro Trp Thr Gly Thr Phe Thr Ser Thr Ser Thr Glu Met Ser Thr Val  
420 425 430

Thr Gly Thr Asn Gly Leu Pro Thr Asp Glu Thr Val Ile Val Val Lys  
435 440 445

Thr Pro Thr Thr Ala Ile Ser Ser Ser Leu Ser Ser Ser Ser Gly  
450 455 460

Gln Ile Thr Ser Ser Ile Thr Ser Ser Arg Pro Ile Ile Thr Pro Phe  
465 470 475 480

Tyr Pro Ser Asn Gly Thr Ser Val Ile Ser Ser Ser Val Ile Ser Ser  
485 490 495

Ser Val Thr Ser Ser Leu Phe Thr Ser Ser Pro Val Ile Ser Ser Ser  
500 505 510

Val Ile Ser Ser Ser Thr Thr Thr Ser Thr Ser Ile Phe Ser Glu Ser  
515 520 525

Ser Lys Ser Ser Val Ile Pro Thr Ser Ser Ser Thr Ser Gly Ser Ser  
530 535 540

Glu Ser Glu Thr Ser Ser Ala Gly Ser Val Ser Ser Ser Ser Phe Ile  
545 550 555 560

Ser Ser Glu Ser Ser Lys Ser Pro Thr Tyr Ser Ser Ser Ser Leu Pro  
565 570 575

Leu Val Thr Ser Ala Thr Thr Ser Gln Glu Thr Ala Ser Ser Leu Pro  
580 585 590

Pro Ala Thr Thr Thr Lys Thr Ser Glu Gln Thr Thr Leu Val Thr Val  
595 600 605

Thr Ser Cys Glu Ser His Val Cys Thr Glu Ser Ile Ser Pro Ala Ile  
610 615 620

Val Ser Thr Ala Thr Val Thr Val Ser Gly Val Thr Thr Glu Tyr Thr  
625 630 635 640

Thr Trp Cys Pro Ile Ser Thr Thr Glu Thr Thr Lys Gln Thr Lys Gly  
645 650 655

Thr Thr Glu Gln Thr Thr Glu Thr Thr Lys Gln Thr Thr Val Val Thr  
660 665 670

Ile Ser Ser Cys Glu Ser Asp Val Cys Ser Lys Thr Ala Ser Pro Ala  
675 680 685

Ile Val Ser Thr Ser Thr Ala Thr Ile Asn Gly Val Thr Thr Glu Tyr  
690 695 700

Thr Thr Trp Cys Pro Ile Ser Thr Thr Glu Ser Arg Gln Gln Thr Thr  
705 710 715 720

Leu Val Thr Val Thr Ser Cys Glu Ser Gly Val Cys Ser Glu Thr Ala  
725 730 735

Ser Pro Ala Ile Val Ser Thr Ala Thr Ala Thr Val Asn Asp Val Val  
740 745 750

Thr Val Tyr Pro Thr Trp Arg Pro Gln Thr Ala Asn Glu Glu Ser Val  
755 760 765

Ser Ser Lys Met Asn Ser Ala Thr Gly Glu Thr Thr Thr Asn Thr Leu  
 770 775 780  
 5 Ala Ala Glu Thr Thr Thr Asn Thr Val Ala Ala Glu Thr Ile Thr Asn  
 785 790 795 800  
 Thr Gly Ala Ala Glu Thr Lys Thr Val Val Thr Ser Ser Leu Ser Arg  
 805 810 815  
 10 Ser Asn His Ala Glu Thr Gln Thr Ala Ser Ala Thr Asp Val Ile Gly  
 820 825 830  
 His Ser Ser Ser Val Val Ser Val Ser Glu Thr Gly Asn Thr Lys Ser  
 835 840 845  
 15 Leu Thr Ser Ser Gly Leu Ser Thr Met Ser Gln Gln Pro Arg Ser Thr  
 850 855 860  
 Pro Ala Ser Ser Met Val Gly Tyr Ser Thr Ala Ser Leu Glu Ile Ser  
 865 870 875 880  
 20 Thr Tyr Ala Gly Ser Ala Thr Ala Tyr Trp Pro Val Val Val  
 885 890

## (2) INFORMATION FOR SEQ ID NO: 16:

25

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs  
 (B) TYPE: nucleic acid  
 30 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPES DNA (genomic)

35

## (vii) IMMEDIATE SOURCE:

(B) CLONE: ChoB template coding strand

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

40

GCCCCCAGCC GCACCCTCG

19

## (2) INFORMATION FOR SEQ ID NO: 17:

45

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs  
 (B) TYPE: nucleic acid  
 50 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

55

## (vii) IMMEDIATE SOURCE:

(B) CLONE: ChoB template non-coding strand



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

**CGAGGGTGCG GCTGGGGGC**

**19**

5

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

15

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

20

(B) CLONE: cho01pcr primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

25

**AGATCTGAAT TCGCGGCCGC CCCCAGCCGC ACCCTCG**

**37**

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

30

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

40

(B) CLONE: cho02pcr primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

45

**AGATCTAAGC TTTCAGCTAG CCTGGATGTC GGAAGAGATG AT**

**42**

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

55

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

(B) CLONE: ChoB template coding strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5

**ATCATCTCGT CCGACATCCA G**

**21**

(2) INFORMATION FOR SEQ ID NO: 21:

10

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

15

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

20

(B) CLONE: ChoB template non-coding strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

25

**CTGGATGTCG GACGAGATGA T**

**21**

(2) INFORMATION FOR SEQ ID NO: 22:

30

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 base pairs

(B) TYPE: nucleic acid

35

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

40

(vii) IMMEDIATE SOURCE:

(B) CLONE: mutagenesis primer ChoB

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

45

**CGCGGCGACG GCACCGCCGT ATGCACTGGC GATGACGAGG GC**

**42**

(2) INFORMATION FOR SEQ ID NO: 23:

50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 base pairs

(B) TYPE: nucleic acid

55

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

(B) CLONE: ChoB template coding strand

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

**CCCCCTCGTCA TCGGCAGTGG ATACGGCGGT GCCGTGCGCG CG**

**42**

10 (2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

15 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

20

(vii) IMMEDIATE SOURCE:

(B) CLONE: primer prtl

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

**AAGATCTATC GATCTTCTTA GCCCGTACA**

**29**

30 (2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

35 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

40

(vii) IMMEDIATE SOURCE:

(B) CLONE: proteinase template non-coding strand

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

**GACTGTACCG GCTAACAAGA TCGATAGCCC TT**

**32**

50 (2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

55 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

5 (B) CLONE: proteinase template coding strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

10 **GTCGGCGAAA TCCAAGCAAA GCGGGCT**

**27**

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

25

(B) CLONE: prt2 primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

30 **CCCAAGCTTC CCCCCGGCCG TTGCTTGGAT TTCGCGGAC**

**39**

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

35

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

40

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

45

(B) CLONE: EGF1 primer

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

50 **GGGGCGGCGG CGCTGGAGGA AAAGAAAGTT TGC**

**33**

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

55

- (A) LENGTH: 33 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

5 (vii) IMMEDIATE SOURCE:

(B) CLONE: EGF receptor template non-coding strand

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

**GCAAACCTTTC TTTCCTCCA GAGCCCGACT CGC**

**33**

15 (2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

20 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

25

(vii) IMMEDIATE SOURCE:

(B) CLONE: EGF receptor template coding strand

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

**AATGGGCCTA AGATCCCGTC CATCGCCACT**

**30**

35 (2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 base pairs

40 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

45

(vii) IMMEDIATE SOURCE:

(B) CLONE: EGF2 primer

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

**CCCCAAGCTT AAGGCTAGCG GACGGGATCT TAGGCCCAT**

**40**

55 (2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 177 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

10

(B) CLONE: VhC - AG $\alpha$ 1 linker with MycT and Hinge

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

15

<b>GAATTCCAGG TCACCGTCTC CTCAGAACAA AAATCATCT CAGAAGAGGA TCTGAATGAA</b>	<b>60</b>
<b>CCAAAGATTG CACAACCTCA ACCAAAGCCA CAACCTCAAC CACAACCACA ACCAAAACCT</b>	<b>120</b>
<b>CAACCAAAGC CAGAACCAGA ATCTACTTCC CCAAAGTCTC CAGCTAGCCT TAAGCTT</b>	<b>177</b>

20

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

25

(A) LENGTH: 63 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

30

(vii) IMMEDIATE SOURCE:

(B) CLONE: VhC - AG $\alpha$ 1 linker with MycT

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

<b>GAATTCCAGG TCACCGTCTC CTCAGAACAA AAATCATCT CAGAAGAGGA TCTGAATGCT</b>	<b>60</b>
<b>AGC</b>	<b>63</b>

40

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

45

(A) LENGTH: 144 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

50

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

(B) CLONE: VhC - AG $\alpha$ 1 linker with Hinge

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

GAATTCAGG TCACGGTCTC CTCAGAACCA AAGATTCCAC AACCTCAACC AAAGCCACAA 60  
 CCTCAACCAC AACCACAACC AAAACCTCAA CCAAAGCCAG AACCAGAATC TACTTCCCCA 120  
 5 AAGTCTCCAG CTAGCCTTAA GCTT 144

(2) INFORMATION FOR SEQ ID NO: 35:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 119 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 15 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

- 20 (B) CLONE: fragment in pUR4421 coding strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

25 AATTTAGCGG CGGCCAGGT GAAACTGCTC GAGTAAGTGA CTAAGGTCAC CGTCTCCTCA 60  
 GAACAAAAC TCATCTCAGA AGAGGATCTG AATTAATGAG AATTCATCAA ACGGTCATA 119

30 (2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 119 base pairs
- 35 (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

40 (vii) IMMEDIATE SOURCE:

- (B) CLONE: fragment in pUR4421 non-coding strand

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

AGCTTATCAC CGTTTGATGA ATTCTCATTA ATTCAGATCC TCTTCTGAGA TGAGTTTTTG 60  
 50 TTCTGAGGAG ACGGTGACCT TAGTCACTTA CTCGAGCAGT TTCACCTGGG CGGCCGCTA 119

55 (2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5

(ii) MOLECULE TYPE: peptide

(vii) IMMEDIATE SOURCE:

10

(B) CLONE: Myc tail

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

15

**Glu Gln Lys Leu Ile Ser Glu Glu Asp Leu Asn**  
**1 5 10**

(2) INFORMATION FOR SEQ ID NO: 38:

20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

30

(B) CLONE: BstEII-HindIII linker coding strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

35

**GTCACCGTCT CCTCATAATG A**

**21**

(2) INFORMATION FOR SEQ ID NO: 39:

40

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

45

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:

50

(B) CLONE: BstEII HindIII linker non-coding strand

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

55

**AGCTTCATTA TGAGGAGACG**

**20**

(2) INFORMATION FOR SEQ ID NO: 40:



## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE: (B) CLONE: primer cho03pcr

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CGGATCCAAG CTTGAGCCTG GATGTCGGAC GAGATGAT

38

## Claims

1. A process for carrying out an isolation process by using an immobilised binding protein or functional part thereof still capable of binding to a specific compound, wherein the immobilised protein is localized at the exterior of the cell wall of a host cell, wherein a medium containing said specific compound is contacted with a host cell under conditions whereby a complex between said specific compound and said immobilised binding protein is formed, separating said complex from the medium originally containing said specific compound, wherein the host cell is a fungus selected from the group consisting of yeasts and moulds, containing an expressible polynucleotide comprising

- (i) a structural gene encoding the binding protein or a functional part thereof still having the specific binding capability, said binding protein or said functional part thereof being localized at the cell wall of said fungus, and  
 (ii) at least a part of a gene encoding an anchoring protein capable of anchoring in the cell wall of said fungus, said part of a gene encoding at least the anchoring part of said anchoring protein, which anchoring part is derivable from the C-terminal part of said anchoring protein;

said polynucleotide being present in a vector or in the chromosome of said fungus.

2. A process according to claim 1, wherein the fungus is selected from the group consisting of yeasts belonging to the genera *Candida*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Pichia* and *Saccharomyces* and moulds belonging to the genera *Aspergillus*, *Penicillium* and *Rhizopus* can be used.
3. Process according to claim 1 or 2 wherein the protein capable of anchoring in the cell wall is selected from the group consisting of, AGA1 (=a-agglutinin) of *S. cerevisiae*, FLO1 (= flocculation protein), Major Cell Wall Protein of lower eukaryotes, selected from the group consisting of yeasts and fungi.
4. Process according to any of claims 1-3 wherein said fungus further comprises a sequence encoding a signal peptide ensuring secretion of the expression product of said first polynucleotide.
5. Process according to any of claims 1-4 wherein the binding protein is an antibody, an antibody fragment, a combination of antibody fragments, a receptor protein, or an inactivated enzyme still capable of binding the corresponding substrate.

## Patentansprüche

1. Verfahren zum Durchführen eines Isolierungsverfahrens durch Verwenden eines immobilisierten Bindungsproteins oder eines funktionalen Teils davon, das (der) immer noch zur Bindung einer spezifischen Verbindung fähig ist, wobei das immobilisierte Protein an der Außenseite der Zellwand einer Wirtszelle lokalisiert ist, wobei ein Medium, das die spezifische Verbindung enthält, mit einer Wirtszelle unter Bedingungen in Kontakt gebracht wird, unter denen ein Komplex zwischen der spezifischen Verbindung und dem immobilisierten Bindungsprotein gebildet wird

und der Komplex von dem die spezifische Verbindung ursprünglich enthaltenden Medium getrennt wird, wobei die Wirtszelle ein Pilz ist, ausgewählt aus der Gruppe, die aus Hefen und Schimmelpilzen besteht, enthaltend ein exprimierbares Polynucleotid, umfassend

- 5 (i) ein Strukturgen, das für das Bindungsprotein oder einen funktionalen Teil davon, der immer noch die spezifische Bindungskapazität aufweist, codiert, wobei das Bindungsprotein oder der funktionale Teil davon auf der Zellwand des Pilzes lokalisiert ist, und
- 10 (ii) zumindest einen Teil eines Genes, das für ein Ankerprotein codiert, das zur Verankerung in der Zellwand des Pilzes fähig ist, wobei der Teil eines Genes zumindest für den Verankerungsteil des Ankerproteins codiert, wobei der Verankerungsteil vom C-terminalen Teil des Ankerproteins ableitbar ist;

wobei das Polynucleotid in einem Vektor oder Chromosom des Pilzes vorliegt.

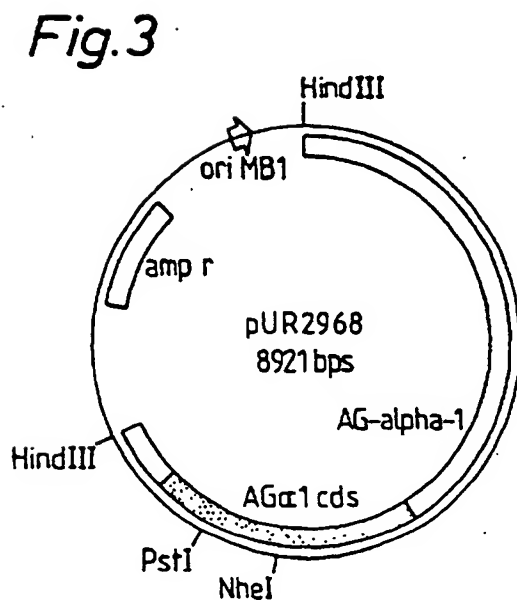
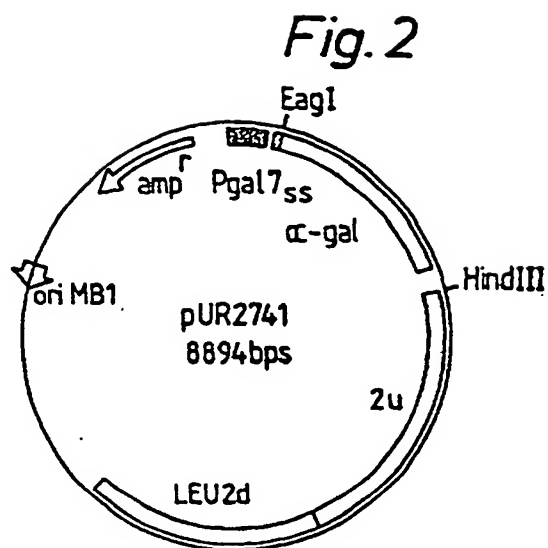
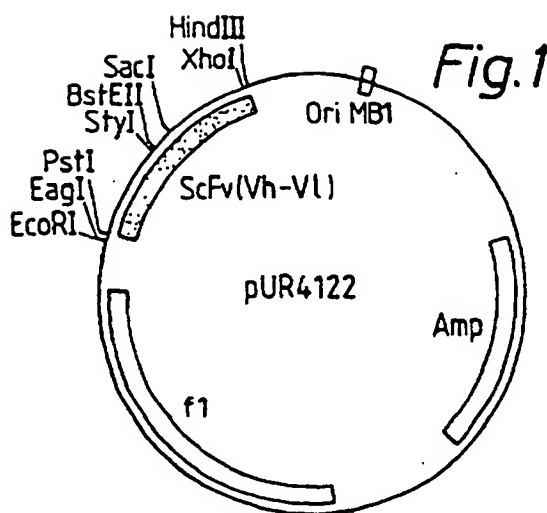
- 15 2. Verfahren gemäß Anspruch 1, wobei der Pilz aus der Gruppe ausgewählt ist, die aus Hefen besteht, die zu den Genera *Candida*, *Debaryomyces*, *Nansenula*, *Kluyveromyces*, *Pichia* und *Saccharomyces* besteht und darüber hinaus Schimmelpilze verwendet werden können, die zu den Genera *Aspergillus*, *Penicillium* und *Rhizopus* gehören.
- 20 3. Verfahren gemäß Anspruch 1 oder 2, wobei das zum Verankern in der Zellwand fähige Proteine aus der Gruppe ausgewählt ist, die besteht aus: AGA1 (=a-Agglutinin) von *S. cerevisiae*, FLO1 (= Flocculationsprotein), Hauptzellwandprotein von niederen Eukaryoten, ausgewählt aus der Gruppe, die aus Hefen und Pilzen besteht.
- 25 4. Verfahren gemäß mindestens einem der Ansprüche 1 bis 3, wobei der Pilz darüber hinaus eine Sequenz umfasst, die für ein Signalpeptid codiert, wodurch die Sekretion des Expressionsproduktes des ersten Polynucleotids sichergestellt wird.
- 30 5. Verfahren gemäß mindestens einem der Ansprüche 1 bis 4, wobei das Bindungsprotein ein Antikörper, ein Antikörperfragment, eine Kombination von Antikörperfragmenten, ein Rezeptorprotein oder ein inaktiviertes Enzym ist, das immer noch zur Bindung des entsprechenden Substrats fähig ist.

## Revendications

- 35 1. Processus adapté pour réaliser un processus d'isolation en utilisant une protéine de liaison immobilisée ou une partie fonctionnelle de celle-ci toujours capable de se lier à un composé spécifique, dans lequel la protéine immobilisée est située à l'extérieur de la paroi cellulaire d'une cellule hôte, dans lequel un milieu contenant ledit composé spécifique est mis en contact avec une cellule hôte dans des conditions où un complexe entre ledit composé spécifique et ladite protéine de liaison immobilisée est formé, séparant ledit complexe du milieu contenant initialement ledit composé spécifique, dans lequel la cellule hôte est un champignon choisi dans le groupe constitué par les levures et les moisissures, contenant un polynucléotide d'expression comprenant
- 40 (i) un gène structurel codant la protéine de liaison ou une partie fonctionnelle de celle-ci ayant toujours la capacité de liaison spécifique, ladite protéine de liaison ou ladite partie fonctionnelle de celle-ci étant située au niveau de la paroi cellulaire dudit champignon, et
- 45 (ii) au moins une partie d'un gène codant une protéine d'ancrage capable de se fixer à la paroi cellulaire dudit champignon, ladite partie d'un gène codant au moins la partie de fixation de ladite protéine d'ancrage, laquelle partie de fixation est susceptible de dériver de la partie C-terminal de ladite protéine d'ancrage ;
- 50 ledit polynucléotide étant présent dans un vecteur ou dans le chromosome dudit champignon.
2. Processus selon la revendication 1 dans lequel il est possible d'utiliser le champignon choisi dans le groupe constitué par les levures appartenant aux genres *Candida*, *Debaryomyces*, *Hansenula*, *Kluyveromyces*, *Pichia* et *Saccharomyces* et les moisissures appartenant aux genres *Aspergillus*, *Penicillium* et *Rhizopus*.
- 55 3. Processus selon la revendication 1 ou 2 dans lequel la protéine capable de se fixer à la paroi cellulaire est choisie dans le groupe constitué par AGA1 (agglutinine A) de *S. cerevisiae*, FLO1 (protéine de floculation), protéine majeure de paroi cellulaire des eucaryotes inférieurs, sélectionnée dans le groupe constitué par les levures et les

champignons.

4. Processus selon l'une quelconque des revendications 1 à 3 dans lequel ledit champignon comprend en outre une séquence encodant un peptide signal garantissant la sécrétion du produit d'expression dudit premier polynucléotide.
5. Processus selon l'une quelconque des revendications 1 à 4 dans lequel la protéine de liaison est un anticorps, un fragment d'anticorps, une combinaison de fragments d'anticorps, une protéine réceptrice ou une enzyme inactivée toujours capable de se lier au substrat correspondant.



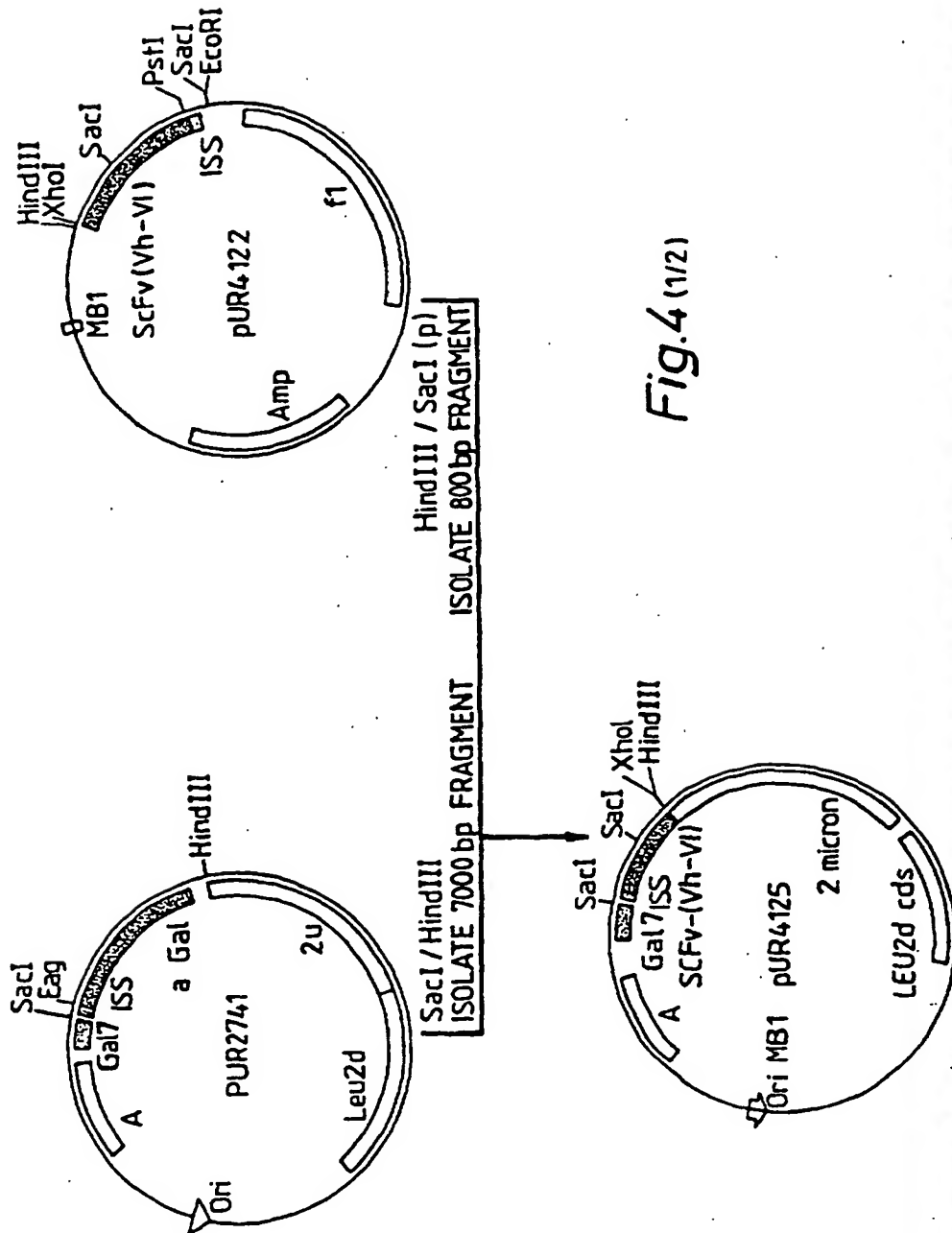


Fig.4 (1/2)

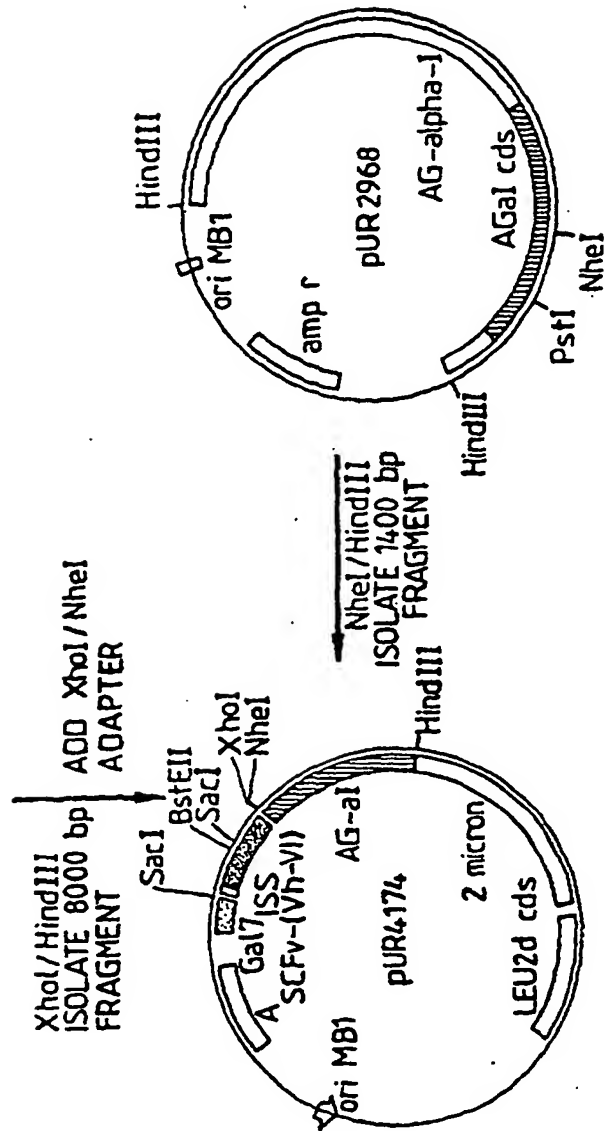


Fig. 4 (1/2) (Cont.)

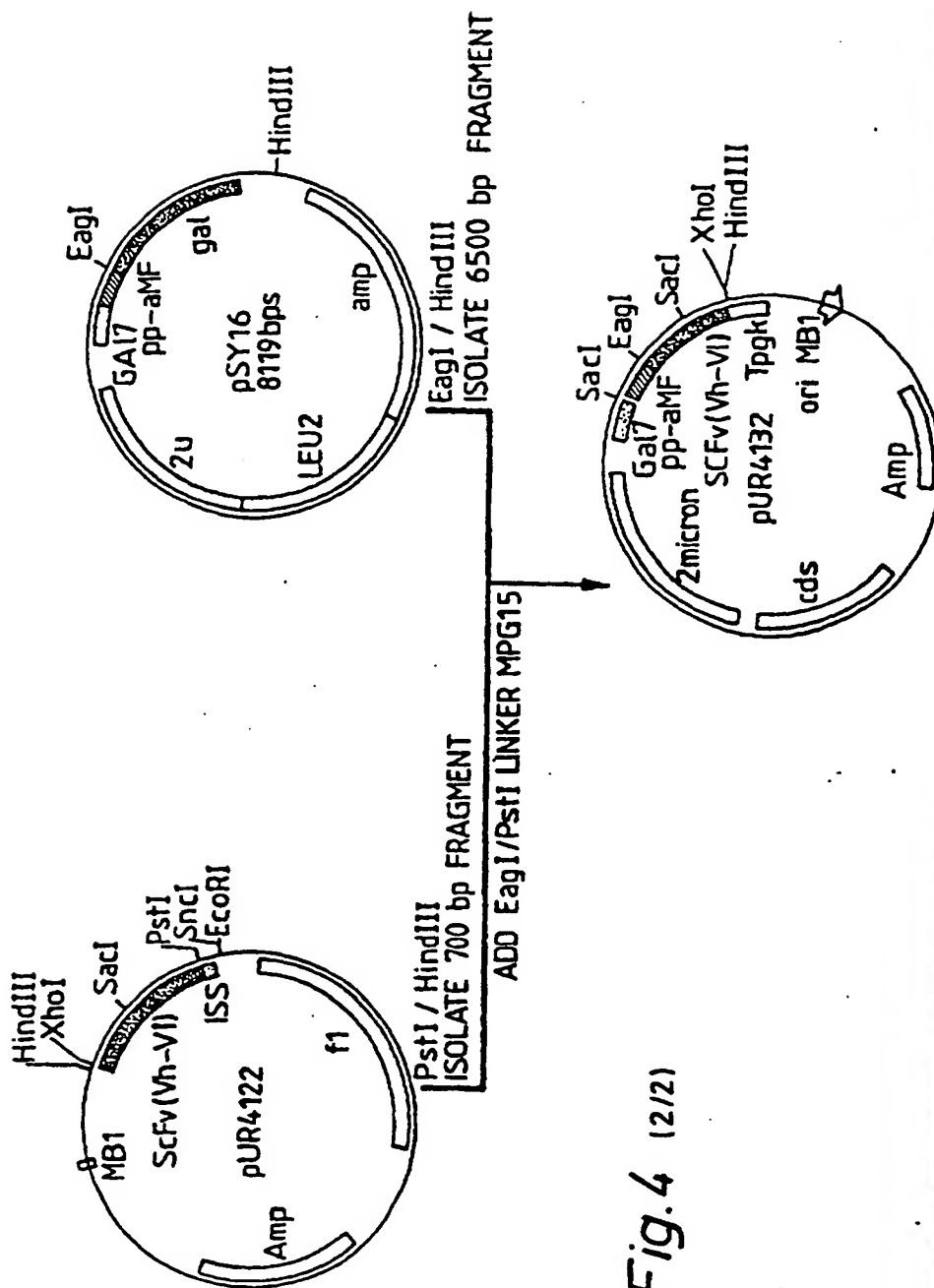


Fig.4 (2/2)

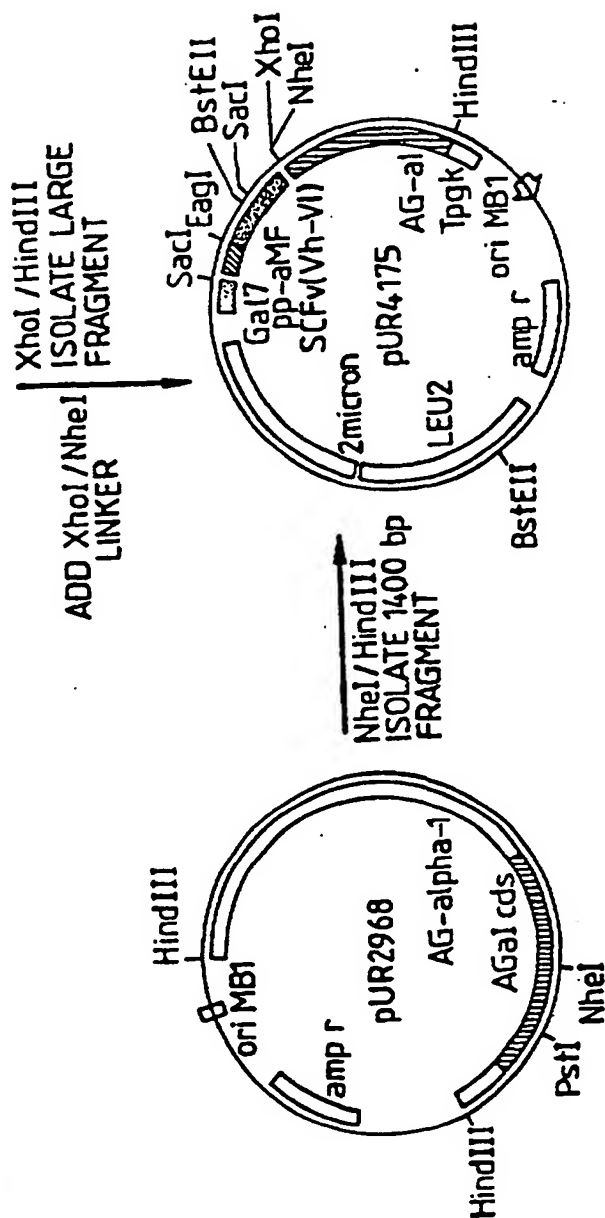
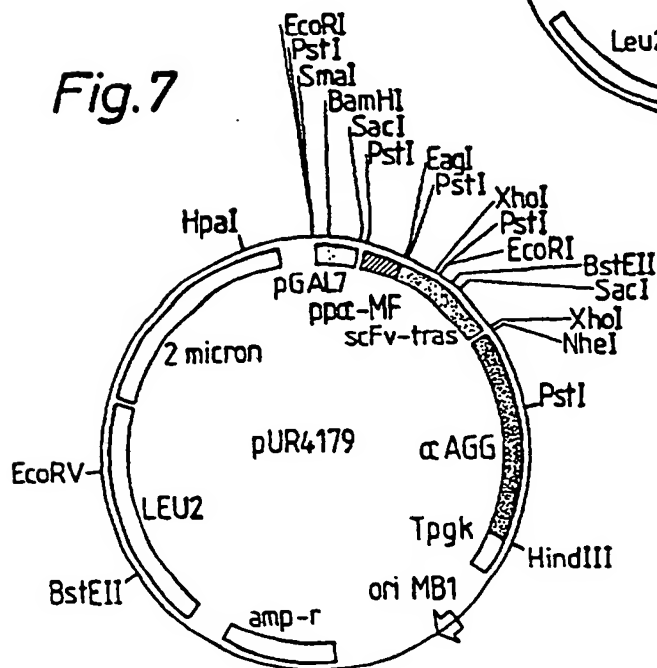
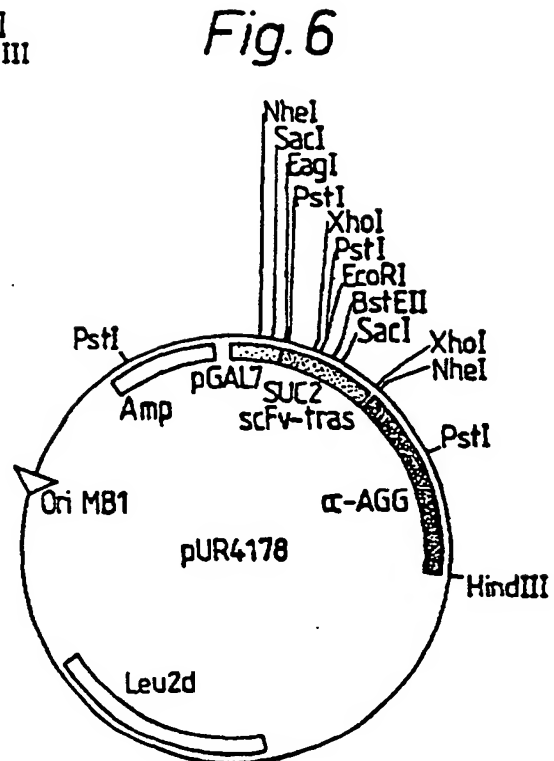
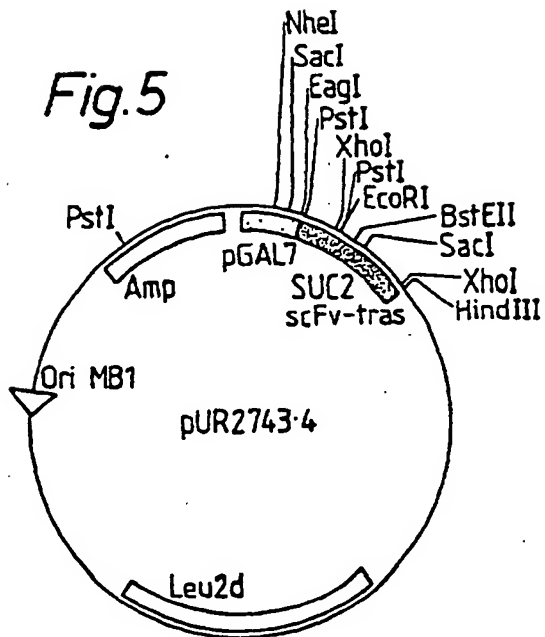
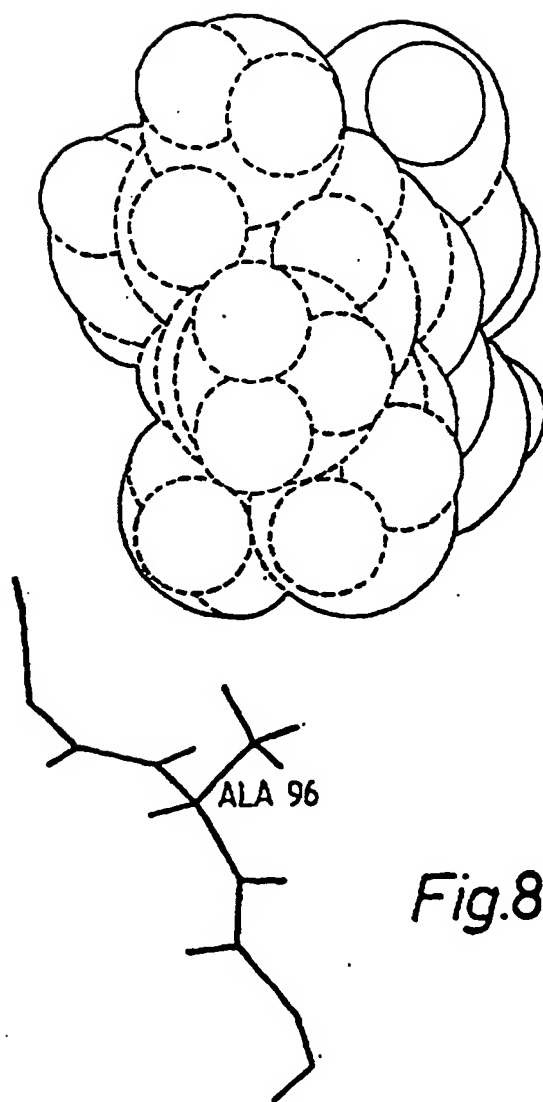


Fig. 4 (2/2) (Cont.)







*Fig.8*

Fig. 9

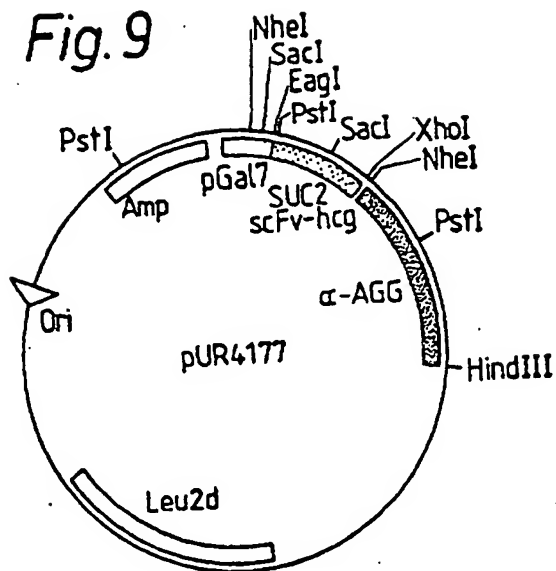


Fig. 10

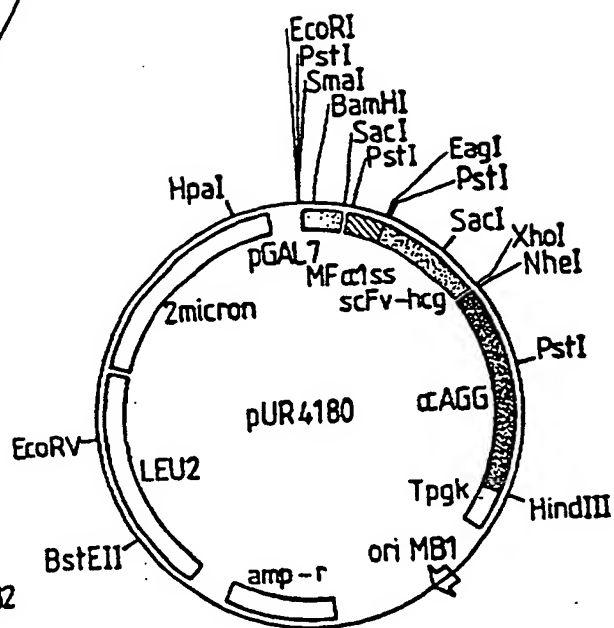
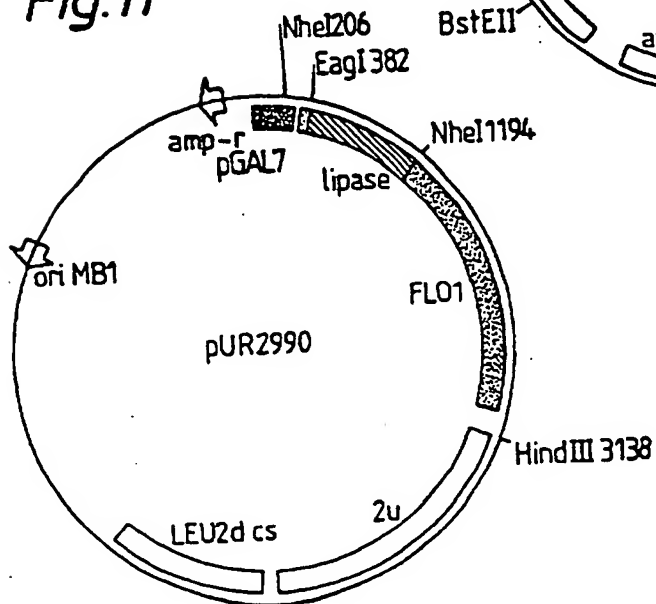
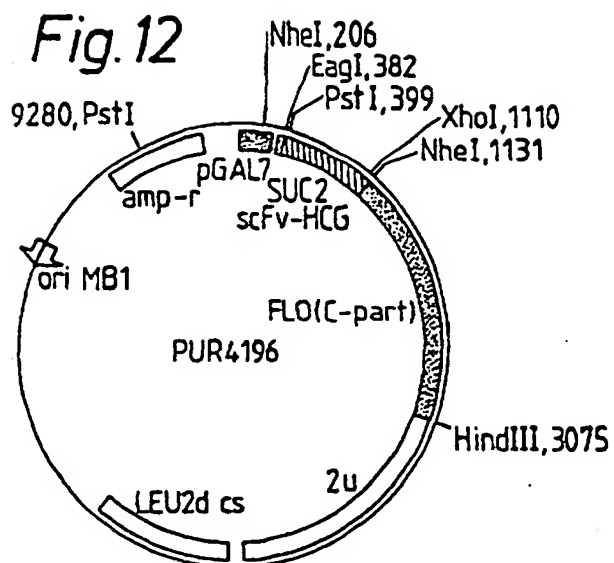
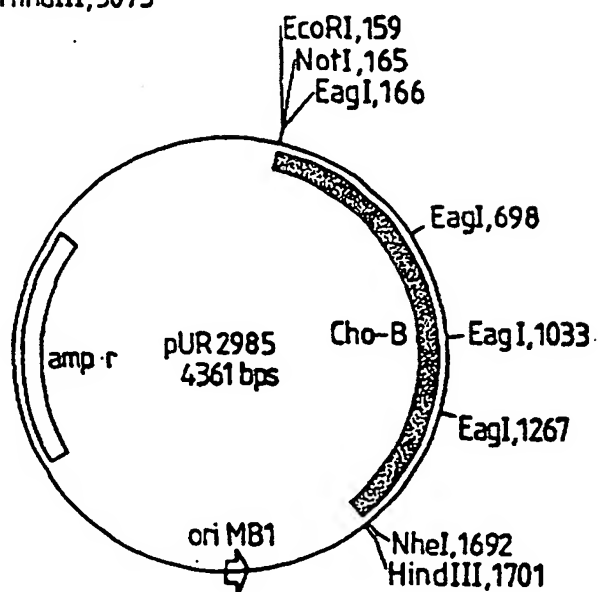


Fig. 11

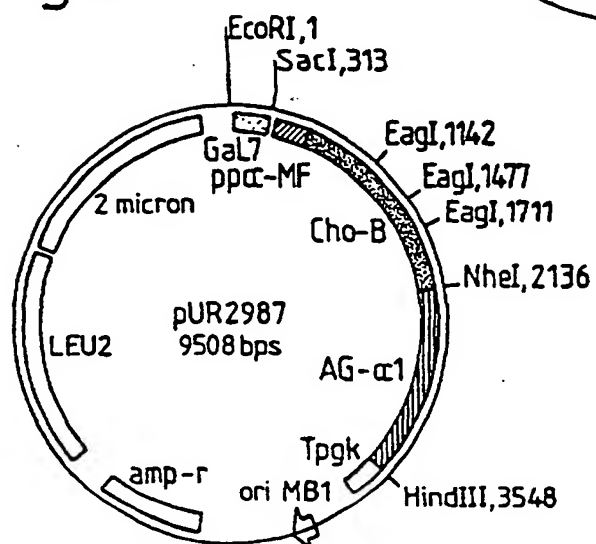


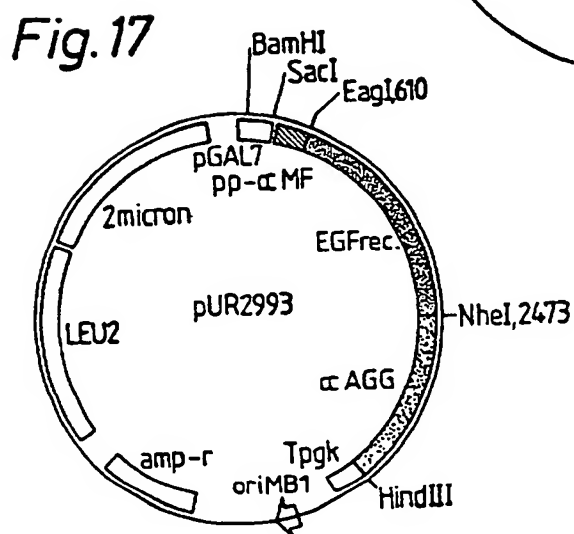
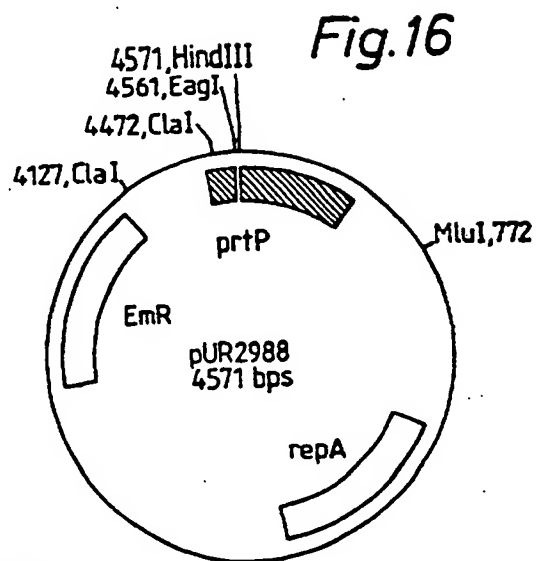
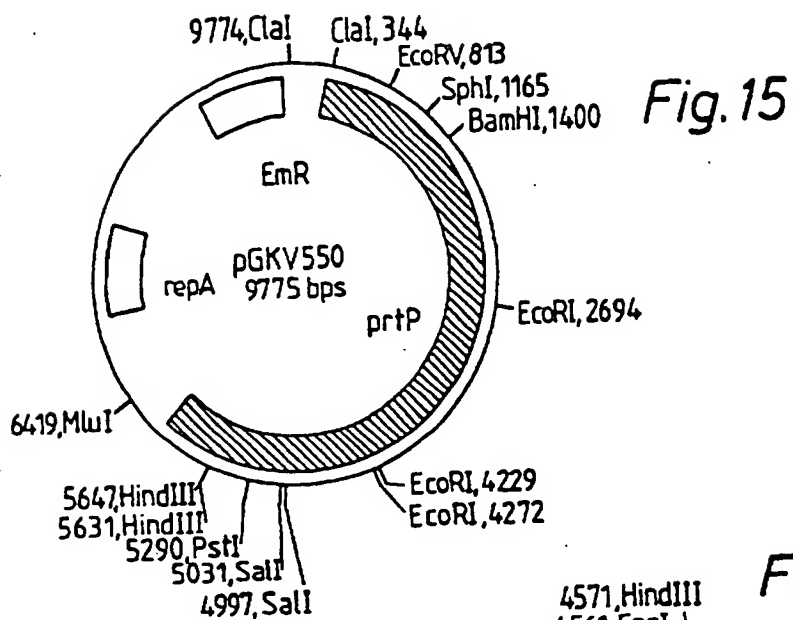


**Fig. 13**



**Fig. 14**





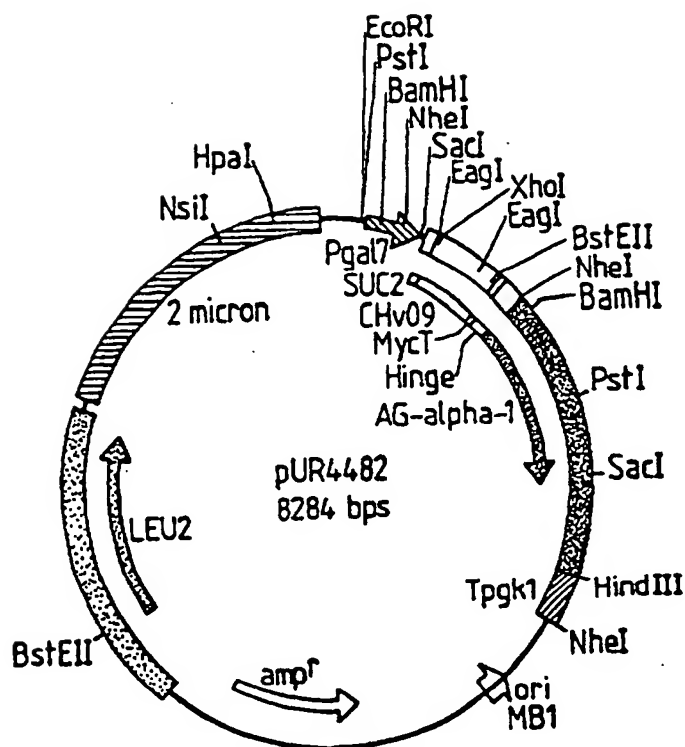


Fig. 18

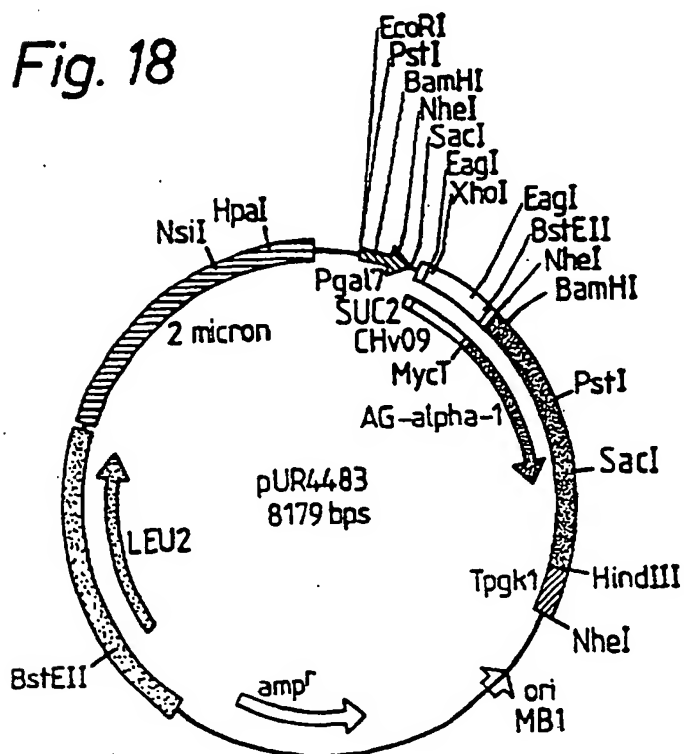


Figure 19

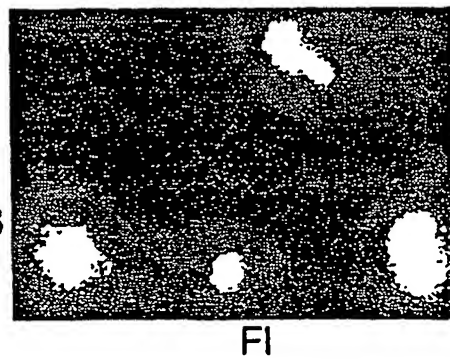
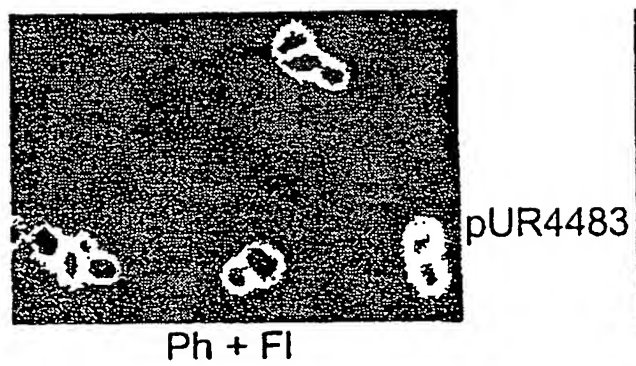
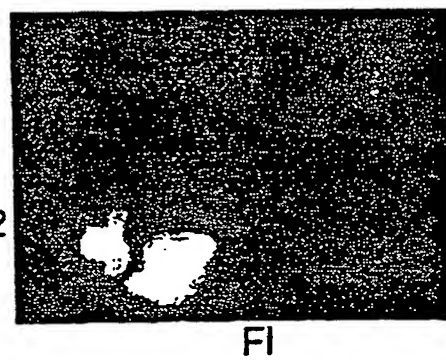
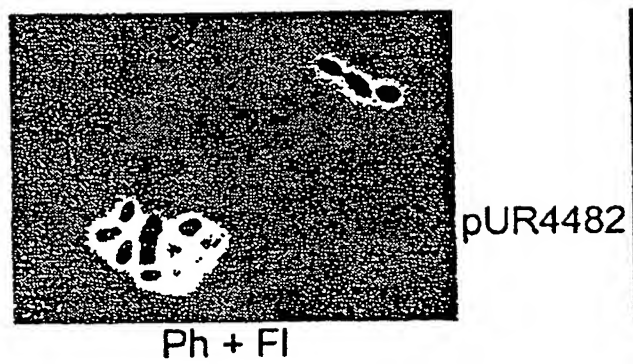
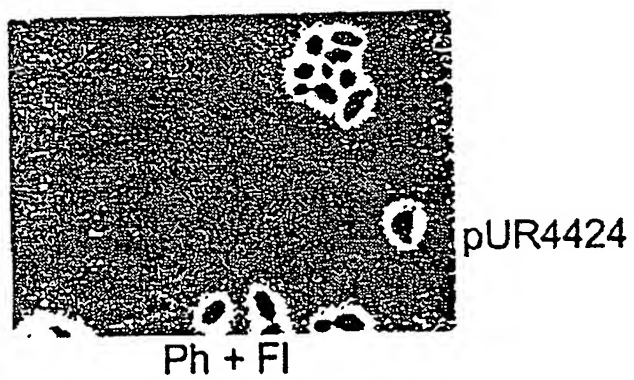
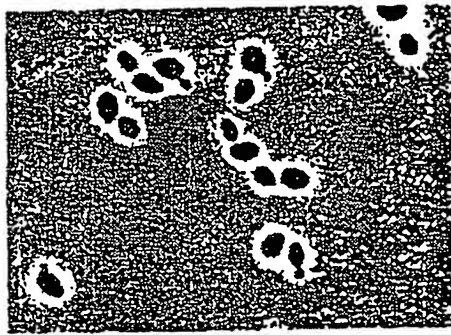
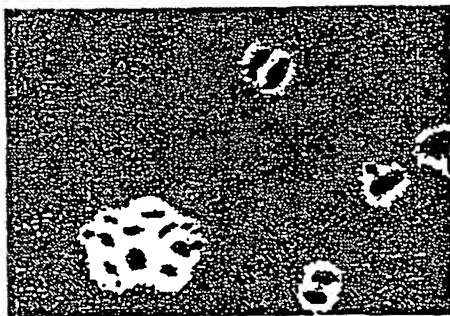


Figure 20



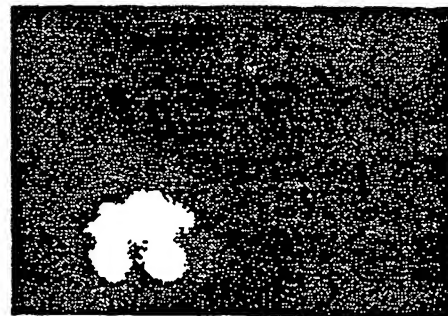
pUR4424

Ph + FI

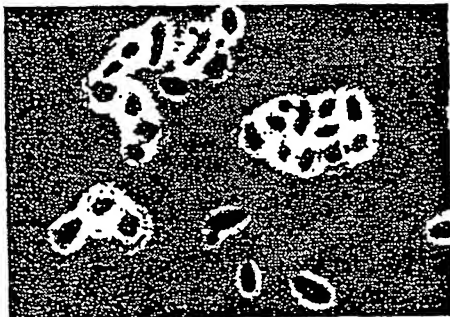


pUR4482

Ph + FI

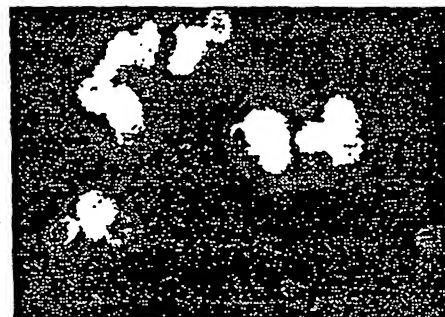


FI



pUR4483

Ph + FI



FI